University Of Alberta

Reliability Estimates and Exploratory Factor Analysis of an American Sign Language Administration of the General Aptitude Test Battery

BY

Denis Cooney ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor Of Philosophy.

in

Special Education

Department of Educational Psychology

Edmonton, Alberta

Spring, 1997

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced with the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-21560-1

Canada

University of Alberta

Library Release Form

Name of Author: Denis Cooney

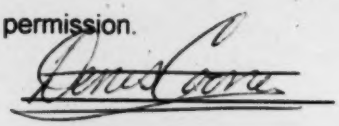
Title of Thesis: Reliability Estimates and Exploratory Factor Analysis of an American Sign Language Administration of the General Aptitude Test Battery

Degree: Doctor of Philosophy

Year Degree Granted: 1997

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form without the authors prior written permission.



Denis Cooney
522 Broadview Avenue
Ottawa, Ontario, K2A-2L4

November 22, 1996

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled RELIABILITY ESTIMATES AND EXPLORATORY FACTOR ANALYSIS OF AN AMERICAN SIGN LANGUAGE ADMINISTRATION OF THE GENERAL APTITUDE TEST BATTERY submitted by DENIS J. COONEY in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Special Education

Harvey W. Zingle, PhD

Michael Rodda, PhD

Jack S. Goldberg, PhD

David J. Sande, PhD

Fern D. Snart, PhD

David G. Mason, PhD

Nov 22 / 96
Date

Abstract

As hearing impaired persons frequently encounter difficulties with the format and procedures of pencil and paper tests with multiple choice items, it was felt that communicating instructions in American Sign language would alleviate most of the difficulties in test administration. From a review of the literature, it was determined that conveying the instructions of the General Aptitude Test Battery in American Sign Language (GATB-ASL) would be appropriate for deaf adults who were fluent in ASL. Standardization was achieved by presenting the instructions on videotape. In addition, animation was used to present answers to practice exercises. In general, the hearing impaired sample experienced little difficulty demonstrating appropriate responses to the instructions conveyed in American Sign Language.

Reliability was estimated by administering the GATB-ASL to 106 deaf and hard-of-hearing adults on two occasions at an interval of six weeks between initial testing and retesting. With the exception of Finger Dexterity, estimates of test-retest reliability were high. The internal consistency (KR-20) of Verbal, Numerical, Spatial, and Perceptual aptitudes was high.

Construct validity was investigated by means of an exploratory factor analysis. Three factors with eigenvalues greater than one were extracted from a correlation matrix and orthogonally rotated to a simple structure. Factor 1 was interpreted as representing a general Cognitive Factor as the measures of Verbal, Numerical, Spatial, and Perceptual aptitudes showed a high loading on Factor 1 and a low loading on the other two factors. Factor 2 represented Manual Dexterity. Factor 3 represented Finger Dexterity

As test booklets, answer sheets, and apparatus boards of the GATB-ASL were identical to the materials used in the standardized version of the GATB, the major deviation from the standardized format and procedures involved presenting instructions on videotape. The effect of the videotape presentation on performance, independent of the pencil and paper multiple choice format, remains unclear.

Table of Contents

Abstract

List of Tables

Chapter Page

1. Introduction 1

Background to the Present Study 2

The Present Study 4

2. Review of The Literature 8

Non-Verbal Methods of Communication 9

Difficulties Administering Tests 14

General Aptitude Test Battery 21

Use with the Hearing Impaired 33

Canadian Edition 38

Summary 39

3. Methods and Procedures 42

Development of GATB-ASL 42

Translation Criteria 42

Translating Procedure 43

Translation Review Procedure 46

Video Production Procedure 47

The GATB-ASL 49

GATB-ASL Administration Procedures 50

Data Collection 52

Sample 52

Scoring 54

Data Analysis 55

Reliability Estimates 55

Factor Analysis 56

4. Results 57

Sample 57

Performance of the Sample 58

GATB-ASL Scores 61

Raw Test Scores 61

Reliability of Test Scores 62

Aptitude Scores 63

Exploratory Factor Analysis 65

Extraction of Factors 66

Final Solution 67

5. Discussion 69

Response to Instructions 70

Reliability Estimates 74

Exploratory Factor Analysis 76

Limitations 78

Conclusion 79

Bibliography 82

Appendix A General Aptitude Test Battery 89

Appendix B GATB-ASL: Typical Questions and Problems 93

Appendix C Computer Print-out of Factor Analysis 97

Appendix D Questionnaire 106

List of Tables

Table		Page
1.	GATB Factor Loadings	24
2.	GATB Correlation Matrix	26
3.	GATB Aptitude Intercorrelations	28
4.	Correlations between Aptitudes and General Factors	28
5.	GATB Test-Retest Reliability Estimates	31
6.	GATB Aptitude Means and Standard Deviations	36
7.	GATB-ASL Sample	58
8.	GATB-ASL Means, Standard Deviations, Mean Differences	62
9.	GATB-ASL Reliability Estimates	63
10.	GATB-ASL Aptitude Means, Standard Deviations, Mean Differences, Reliability	65
11.	GATB-ASL Correlation Matrix	66
12.	GATB-ASL Unrotated Factor Matrix	67
13.	GATB-ASL Rotated (Orthogonal) Factor Matrix	68

Chapter 1

Introduction

The General Aptitude Test Battery (GATB) has been administered to deaf and hard-of-hearing adults on several occasions, but standard procedures for conveying the instructions of the GATB to deaf and hard-of-hearing persons have never been reported in the literature. Therefore, the primary purpose of this study was the development of standard procedures for communicating the instructions of the GATB to deaf and hard-of-hearing adults.

From a review of the literature, it was determined that conveying the directions of the General Aptitude Test Battery (GATB) in American Sign Language (ASL) would be appropriate for deaf and hard-of-hearing persons who are fluent in ASL. Standardization was achieved by presenting the instructions on videotape. In addition to sign language, animation was used to show the correct answers to some of the practice exercises.

Empirical evidence in support of the argument that the directions conveyed in ASL have the same clarity, emphasis, and detail as directions conveyed in the English language came from two main sources. One source of evidence was a comparison of the instructions given in ASL with the instructions given in the English language.

A comparison of a translation with an original text provides evidence for the accuracy, clarity and emphasis of the translation; however, such a comparison does not show that test takers have been provided with sufficient

detail so as to be able to replicate the intentions of the test developer. To determine whether or not instructions conveyed in ASL provided sufficient detail, observations of the performance of test subjects were recorded for each administration of the GATB-ASL.

As the GATB was developed for a hearing population, developing the GATB-ASL for deaf persons gave rise to questions about its reliability and validity. Therefore, the present study also provides evidence of the reliability and the validity of the measures yielded by the GATB-ASL.

Background to the Present Study

Early in this century, Pintner (1915) demonstrated the inappropriateness of the English language in measuring the intelligence of deaf children. Over the next fifty years (1916-1966), researchers studied the validity of performance scales as measures of intelligence, and they compared deaf children with hearing children on numerous scales. It was during this time that pantomime and demonstration were used to standardize the administration of the Hiskey-Nebraska Learning Aptitude Scales (Hiskey, 1966). In terms of non-verbal methods of communication (pantomime, demonstration, and the various methods of signing) the conventional wisdom was that the different methods of non-verbal communication were deemed to have little effect on the scores of deaf children who successfully demonstrated mastery of test taking tasks.

As a consequence, pantomime, demonstration, gestures, finger spelling, and sign language were often used interchangeably in the testing of deaf children.

About fifteen years ago, a few researchers began suggesting that deaf children often responded differently to the different methods of non-verbal communication. In other words, the different methods of non-verbal communication may not have been as interchangeable as was once thought. Within a few years of the public recognition of American Sign Language (ASL), researchers began reporting statistically significant differences between mean scores obtained with pantomime, demonstration, and American Sign Language. Morgan and Vernon (1994) have affirmed that most psychologists now recognize the superiority of using American Sign Language to test deaf persons who are fluent in American Sign Language.

As programs and services for deaf and hard-of-hearing persons have traditionally been under funded, a typical assessment practice among professionals with deaf and hard-of-hearing clients has been to modify tests which had been standardized with a hearing population. One of the difficulties with this procedure is that modifications to standardized testing to accommodate deafness have often been shown to be inappropriate for deaf persons. As the literature on testing deaf persons is meager, only a few researchers have noted the difficulties encountered in the administration of standardized tests to deaf persons.

Trybus (1973) suggested that the difficulties encountered in the administration of standardized tests to deaf persons relate to two main factors. One of the factors is the format and the procedures of pencil and paper tests with multiple choice items. The other factor is the English language reading grade level of test items. Bragman (1982a) suggested that these two factors must also be considered in light of the test taker's familiarity with the non-verbal method of communicating test instructions.

In summary, the difficulties administering standardized tests to deaf persons involve the format and procedures of pencil and paper tests with multiple choice item, and the English language reading grade level of items. Based on Trybus (1973), some of the difficulties associated with the format and procedures of paper and pencil tests may be alleviated by giving test directions in a method of communication with which examinees are familiar.

The Present Study

The primary purpose of the present study was the development of uniform procedures that were logically appropriate for communicating the instructions of the General Aptitude Test Battery (GATB) to deaf and hard-of-hearing adults.

A review of the literature on administering tests to deaf persons indicated that communicating the instructions of the General Aptitude Test Battery in American Sign Language (GATB-ASL) would be appropriate for deaf adults

who are fluent in ASL. Standardization of test administration was achieved by presenting all of the instructions on videotape. In addition to ASL, animation and graphics were used to show the correct responses to practice exercises.

The effect of an accurate ASL interpretation should be that directions conveyed in ASL are as clear and as easy to understand as directions conveyed verbally. Although verbal and non-verbal methods of communication would share similar characteristics (clarity and effortlessness of understanding) they would not be interchangeable. If two methods of communication are not interchangeable, then the obvious questions are how do the methods differ and is the difference meaningful.

The immediate challenge was that a simple comparative study was not practicable. What was feasible was developing standard testing procedures, conducting a test-retest reliability study, and conducting an exploratory factor analysis of the data.

Reliability estimates were computed with the data yielded by administering the GATB-ASL to 106 deaf and hard-of-hearing adults on two occasions with an interval of six weeks between initial testing and retesting. Test-retest reliability was estimated by the Pearson Product Moment Correlation between the scores. The Kuder-Richardson Formula Number 20 (KR-20) was used to estimate the internal consistency of the seven paper and pencil tests with multiple choice items. The KR-20 was not applied to the five performance

tests as each result was the number of times single tasks were performed correctly.

The goal of data reduction was achieved by representing the largest amount of covariance among the twelve tests of the battery with the smallest number of factors. Factors with eigenvalues greater than one were extracted from a correlation matrix and orthogonally rotated to a simple structure. Tests with a high loading on the same factor and a correspondingly low loading on the other factors were assumed to have large amounts of shared variance. With respect to the construct validity of the GATB-ASL, it was assumed that tests measuring different constructs would show low correlations with each other, and tests measuring the same construct would show high correlations with each other.

In summary, the literature on the General Aptitude Test Battery (GATB) indicates that the GATB has been administered to deaf and hard-of-hearing persons on several occasions (the United States Employment Service, 1970), but evidence in support of the reliability and validity of the scores yielded by the GATB have not been reported in the literature (Botterbusch and Droege, 1972). Given that some of the difficulties encountered in the administration of standardized tests to deaf persons involve: the format and procedures of pencil and paper tests, and the English language reading grade level of the

items, the reliability and validity of the GATB scores of deaf persons were opened to question.

The present study involved administering the GATB-ASL to 106 deaf and hard-of-hearing adults on two occasions with an interval of six weeks between initial testing and retesting. Internal consistency was computed with Kuder Richardson Formula Number 20. The test-retest reliability estimate was the Pearson Product Moment Correlation Coefficient. The construct validity of the GATB-ASL was investigated by means of an exploratory factor analysis. The factor analysis extracted three factors with eigenvalues greater than one from a correlation matrix. The three factors were orthogonally rotated to a simple structure.

Chapter 2

Review of the Literature

Many researchers (Bragman, 1982a; Chovan and Benfield, 1994; Holm, 1987, Levine, 1974; Morgan and Vernon, 1994; Sanderson, 1974; Watson, 1979) have noted the scarcity of studies on administering tests to deaf adults. As a consequence, papers on the testing of deaf adults frequently draw on clinical and educational work with deaf children. The literature search for the present study was no exception.

This review of the literature is organized around the works of two major authors. Levine (1974) provides descriptive data on the frequency with which the non-verbal methods of communication are used in testing, the types of difficulties encountered in the testing of deaf children and adults, and the types of test used most often with deaf and hard-of-hearing persons. Vernon (1968) describes the appropriateness of specific tests in the assessment of deaf children, and defends the superiority of sign language over other methods of communication in the testing of deaf persons (Morgan and Vernon, 1994). References to the research of other authors usually provide a description of the performance of deaf and hard-of-hearing persons under specific test conditions.

An attempt was made to organize this review of the literature under three main headings: non-verbal methods of communication, difficulties administering tests, and the General Aptitude Test Battery (GATB). The attempt was only partly

successful as the methods of communication and the difficulties administering tests overlap. Before delving into the literature on the GATB, the research on the non-verbal methods of communication in testing, and the difficulties encountered in the testing of deaf children and adults is reviewed.

Non-Verbal Methods of Communication

Pintner (1924) noted that early in the history of psychological testing the feud between the oralists, advocates of speaking, and the manualists, advocates of signing and finger-spelling, often precluded systematic study of the effects of communication on the measurement of the intelligence of deaf children. As a matter of fact, Pintner affirmed that he and Patterson (Pintner, 1924) were the first to undertake a systematic investigation of the intelligence of deaf children. They attempted to communicate the instructions of the Goddard Revision of the Binet Scale to 22 deaf children by various methods, such as, writing, speech, manual spelling, signs, and / or, combinations of these methods. Their study demonstrated the inappropriateness of measuring the intelligence of deaf children with scales that involve the use of the English language. Pintner suggested that scales involving the use of the English language "immediately become for the deaf a subject-matter test". He emphasized that such tests measured the language deprivation of deaf children rather than their intelligence. To obtain an estimate of the intelligence of deaf children that was independent of language,

Pintner strongly recommended the use of non-language administrations of non-verbal performance tests.

Vernon's (1968) review of testing hearing impaired children over the next fifty years (1917-1967) indicated that the emphasis of research was on validating performance scales as measures of intelligence, and comparing the performance of deaf children with that of hearing children on various scales. It was during this period that Hiskey (1966) used pantomime and demonstration to standardize the administration of his learning aptitude scales. By standardizing procedures that controlled for the language variable, Hiskey was able to develop norms for deaf children and hearing children on the same scales. In describing the development of his learning aptitude scales, Hiskey (1966) defended his strong conviction that any deviation from standardized procedures affects the performance of examinees which, in theory, affects the validity of the inferences made from the scores. Goetzinger and Rousey (1957) were guided by the same rationale in their use of pantomime and demonstration to standardize the Performance Scale of the Wechsler Intelligence Scale Children.

Levine (1974) provided one of the first accounts of the frequency with which different methods of non-verbal communication were used in the psychological testing of hearing impaired persons. Educational facilities and agencies which served a combined clientele of some 24,224 hearing impaired persons responded to Levine's survey. Approximately 52% of the 162 respondents reported using

various combinations of speech, writing, gesture, pantomime and interpreters to test hearing impaired persons. Approximately 31% of the respondents used various combinations of signs, fingerspelling, gesture, pantomime, drawing and speech. The use of gesture and pantomime was reported by 5% of the respondents. The use of sign language Interpreters was reported by only 3% of the respondents. Writing was used in combination with gesture by 3% of the respondents. Fingerspelling and speech were used by another 3% of the sample; and 2% of the respondents used writing as the sole means of testing.

In discussing non-verbal methods of communicating test directions, Sanderson (1974) expressed the opinion that as long as deaf students performed test taking tasks correctly, non-verbal methods of communication should be viewed as having only a minor influence on performance. Sanderson reported, however, that the amount of time required by students to "catch-on" to the demands being made of them frequently depended on the non-verbal method of communication that was used to convey directions. Sanderson observed that some students received instructions by several methods of communication before demonstrating mastery of test taking tasks, such as marking answers in the correct spaces on separate answer sheets. Sanderson also noted that it was only after completing two or three tests that many deaf persons began to demonstrate mastery of the test taking tasks associated with the format and procedures of pencil and paper tests with multiple items.

Sullivan and Vernon (1979) found that administrations of Wechsler's Performance Scale to deaf children who used Total Communication resulted in significantly higher scores than pantomime and demonstration administrations. Sullivan and Vernon concluded that psychological evaluations of deaf children who know sign language should be done in sign language. Sullivan and Vernon noted that some 90% of the professionals who administered tests to hearing impaired children were unable to communicate in either Total Communication or sign language. They questioned the validity of the inferences made from test scores that were obtained with the assistance of sign language interpreters because they felt that by depending on the use of interpreters psychologists lose rapport with clients, as well as, insights that come from direct inter-personal communication.

Murphy and Fleischer (1977) reported that high school material interpreted to deaf students in American Sign Language (ASL) resulted in significantly higher test scores than material interpreted in Signed English (Siglish). When they conducted a similar study with college level material, however, they failed to confirm the superiority of either sign language system (Murphy and Fleischer, 1977). Murphy and Fleischer (1977) attempted to control for the sign language preference (ASL and Siglish) of students at California State University, Northbridge. A sample of twenty-nine students (better ear hearing loss of greater than 80db) was divided into preferred language groups. The two groups were sub-divided so that half received a brief lecture and short test in their non-

preferred language. In each situation, audio-taped lectures were interpreted in a manner similar to a normal classroom setting. The content of the lectures was carefully selected to insure normal delivery of the lecture. Interpreters carefully rehearsed their presentations to insure strict adherence to the syntax of each language system. The results of the study showed no statistically significant difference in test scores. Murphy and Fleischer reported that the study failed to confirm the superiority of either sign language system in communicating test questions to profoundly deaf young adults. They concluded that their sample functioned as a bilingual group, that is, the students performed the same tasks in the same manner independent of sign language preference.

Morgan and Vernon (1994) encapsulated the research on communicating directions to deaf and hard-of-hearing children and adults. They recognized Total Communication and American Sign Language as the preferred methods of communication among most deaf and hard-of-hearing persons in North America. Further, they have argued that research has demonstrated the superiority of sign language over other methods of communicating test instructions to deaf children and adults.

Morgan and Vernon, affirmed that the receptive and expressive English language skills of the average prelingually deaf student at age 17 are equivalent to a 4.0 reading grade level; and that these levels tend to remain constant throughout adulthood. Thus, items and instructions presented in written format

"must be presented at the reading level of test takers, or else given in sign language". Finally, Morgan and Vernon recognized that as "assessment tools and procedures need to be adapted and modified to meet the needs of each individual who is tested", the essential questions will always involve the validity and reliability of scores obtained under modified test condition.

In summary, the use of the English language in measures other than English language achievement has been shown to be inappropriate with hearing impaired persons. Test directions should be conveyed to profoundly deaf test takers by a non-verbal method of communication with which the deaf person is familiar. The practice of using pantomime, demonstration, and sign language interchangeably has been shown to be open to question. The superiority of sign language over other methods of manual communication, such as pantomime and demonstration, has been demonstrated with deaf children (Sullivan and Vernon, 1979). The superiority of ASL over Siglish as a means of communicating test instructions, however, has not been demonstrated with college level students.

Difficulties Administering Tests

Levine's (1974) survey identified the frequency with which different types of difficulties were encountered in the psychological testing of hearing impaired persons. Approximately 46% of the respondents cited the lack of tests, the absence of norms, problems of interpretation and assessment, and problems of test selection and administration as major problems. Difficulties in communicating

were reported by 37% of the respondents. Approximately 9% of the respondents reported difficulties with the behavior of test subjects. The lack of psychologists who were trained to assess deaf clients was reported by 7% of the respondents. The major difficulties reported by 1% of the sample involved inadequate testing facilities and evaluations involving multiplehandicaps. Approximately 4% of the sample reported no difficulty testing hearing impaired persons.

Trybus (1973) attributed the difficulties encountered in the administration of standardized tests to deaf adults to two main factors. One of the factors was the deaf person's lack of familiarity with the format and procedures of pencil and paper tests with multiple choice items. The test taking tasks of locating items in test booklets, and marking answers on separate answer sheets were identified as being particularly troublesome during timed tests. The other factor was the English language reading grade level of test items.

Bragman (1982b) argued that the difficulties encountered with the format and procedures of testing deaf children may involve a lack of familiarity with the method of communication that is used in testing. Bragman found that the difficulties associated with figuring-out how to respond to unfamiliar formats and procedures were magnified by the use of unfamiliar methods of communication. Bragman's study is important because it is one of the few studies that controlled for the manual communication variable, as well as, the item presentation format variable.

Bragman's (1982b) study of pantomime, demonstration and simultaneous communication (speech plus a variation of signed English) indicated that these methods may not be interchangeable in the testing of prelingual deaf children. Bragman administered a pattern recognition test to prelingual deaf children who were 6 years to 8 years of age. The test consisted of a series of model sequence pattern cards and twenty corresponding choice cards. The children were shown the model sequence pattern cards which they were supposed to match with selections from the choice cards. The children were assigned to one of three groups, each group represented one non-verbal method of communication. In addition, each child within a group was randomly assigned to one of two orders of task presentation which corresponded to (1) identical followed by reverse pattern recognition or (2) reverse followed by identical pattern recognition. From her analysis of the data and observations of the children during testing, Bragman concluded that when the task presentation (identical recognition) and method of communication were familiar to the children, performance was similar for the demonstration method, pantomime method, and simultaneous communication. When task presentation (reverse recognition) and the method of communication were not familiar to the children, performance was similar for pantomime method and simultaneous communication method. However, the performance of the children was significantly changed for the demonstration method. Bragman observed that regardless of the method of communication, 78% of the incorrect responses to the reverse pattern recognition task were the correct responses for

the identical pattern recognition task, independent of the order of task presentation. Bragman concluded that prelingual deaf children usually performed tasks in a manner consistent with the intent of test developers when the method of conveying directions and the task were familiar to the children; but the children did not perform tasks in the same consistent manner when the task and the method of conveying directions were unfamiliar to them.

Bragman expressed the opinion that each of the methods of communication represented a different test condition. She suggested that simultaneous communication demanded that the children recognize signs and understand their meaning before attempting a task. Pantomime focused attention on unfamiliar hand movements which had to be figured-out before the task could be attempted. The demonstration method demanded that the children reproduce (mirror image) the demonstration in order to understand the task. Bragman noted that selecting the most appropriate method of communicating directions to a prelingually deaf child is a complex and difficult undertaking. She has suggested that test administrators may be required to use more than one non-verbal method of communication to convey directions to any given prelingually deaf child. Further, as non-verbal methods of communication may not be equivalent, each method may have to be standardized.

Sullivan and Vernon (1979) noted that pencil and paper tests with multiple choice items have been considered to be suitable for testing hearing impaired

youths with well developed expressive and receptive language skills. With respect to testing the English language vocabulary of hearing impaired youths, Sullivan and Vernon (1979) noted that many of the signs which were used to differentiate between choices of the Peabody Picture Vocabulary Test were not visually or cheremically (i.e., discrete gestures that convey meaning) similar. They suggested that using signs with slightly dissimilar meanings to test English language vocabulary often resulted in scores that reflected "associative rather than receptive vocabulary". Sullivan and Vernon (1979) also cited Rosen's (1967) study which indicated that deaf students with a reading grade level comparable to the reading grade level required by the MMPI frequently did not understand test items with idiomatic expressions.

In replicating Miller's (1984) study of the mediating effects of three methods of manual communication on the performance of profoundly deaf students on the WISC-R Verbal subtests, Chovan and Benfield (1994) reported that Verbal IQ scores were approximately 1 to 1.5 standard deviations below the mean Verbal IQ of 96. Although low scores on Verbal sub-tests significantly correlated with Verbal IQs, none of the Verbal sub-test scores significantly correlated with Performance IQs. Chovan and Benfield suggested that their data do not support Miller's report of normal Verbal IQs of profoundly deaf students and question the mediating effects of ASL, Siglish, and Pidgin Sign English, on measurements of the English language vocabulary of profoundly deaf students.

A comparison of a paper and pencil test with multiple choice items with an Ameslan videotape presentation of the same items failed to demonstrate the interchangeability of the two testing formats. Dwyer and Wincenciak (1977) translated three factors of the 16 P. F. Form E into Ameslan and recorded the translation on videotape. The pencil and paper test and the videotape test were administered to a small sample of profoundly deaf college students whose English language reading grade level fell between 3.0 and 6.0 on the *Gates Basic Reading Survey*. All of the students were fluent in Ameslan. The authors noted that all of the students scored significantly higher on the Ameslan videotape presentation of only one of the three factors, and that scores on the other two factors were unaffected by the test format. From their analysis of the data, the authors concluded that the Ameslan videotape format and the English language pencil and paper format were not interchangeable formats for testing profoundly deaf young adults.

Although videotaping is a relatively recent innovation in testing, Shiels' (1980) review of videotaped instructions for Valpar Work Samples concluded that the use of videotaped instructions "in spite of drawbacks, seems to offer more potential as a method of standardizing administration procedures for deaf subjects than do other methods". Shiels' concerns focused on the development of a videotape presentation "that would not be above the minimal-language-skills deaf person and at the same time not alienate the fluent, higher-verbal-signer".

Shiels suggested that one way of solving the problem would be to produce videotapes at different levels of sign language difficulty. Shiels' noted, however, that the cost of producing videotapes and establishing norms for several levels of sign language difficulty would be prohibitive.

In summary, the major difficulties encountered in the administration of tests to hearing impaired persons involve pencil and paper tests with multiple choice items, and items with inappropriate English language reading grade levels. When deaf children are familiar with the test and the method of communication, their performance is appropriately consistent. When deaf children read at a level below the level of test items, the items are inappropriate. The troublesome aspects of pencil and paper tests with multiple choice items consist of tasks such as locating the items in test booklets, and marking answers in the proper spaces on separate answer sheets. A comparison of an English language pencil and paper presentation of test items with a videotaped presentation of the same items conveyed in Ameslan indicates that the two methods of administering a test to profoundly deaf adults may not be interchangeable. Finally, the notion that verbal items interpreted in sign language yield valid test scores has not been supported by research.

General Aptitude Test Battery

Between 1942 and 1944, the United States Employment Service (the USES) developed the General Aptitude Test Battery (GATB) by factor analyzing some 59 employment aptitude tests. Approximately, 95% of the test subjects were white males enrolled in national defense training courses. The age of the subjects ranged from 17 years to 39 years, with a mean of 29 years. All of the subjects completed, at least, 6 years of education. To conduct comparative studies, the sample was sub-divided into eight groups. With the data collected from the eight groups, the USES conducted nine factor analysis studies.

The USES (1970) reported employing Thurstone's methods of multiple-factor-analysis to extract the centroid factors from correlations among the 59 tests and rotating them to a meaningful structure of underlying aptitudes. The solutions in all eight groups were orthogonal simple structures. The USES (1970) reported that, "in as much as the solutions were very nearly orthogonal, and not so exact that different investigators would have obtained identical correlations between factors, it was decided to impose an orthogonal structure on each group and the rotational process was continued until this was achieved". The "smallest number of common factors established in any group was seven, and the largest was ten" (the USES, 1970). In all, the USES identified eleven factors which were identified as: (G) Intelligence, (V) Verbal, (N) Numerical, (S) Spatial, (P) Form Perception,

(Q) Clerical, (A) Aiming, (T) Motor Speed, (F) Finger Dexterity, (M) Manual Dexterity, and (L) Logic.

The USES did not pursue the development of tests that measured Factor L because Factor L was limited to only two of the eight groups and only one test. The USES reported uncertainty interpreting Factor G. Factor G was found in each one of the nine studies and was present in about two dozen tests. The USES (1970) noted that Factor G, "like all of the other factors, is an independent first-order factor established in a position orthogonal to all the rest". The USES noted that almost all of the Numerical tests and all of the Verbal and Spatial tests showed significant projections on Factor G. Factor G was also present in a letter series test, a word memory test, and a perceptual relations test, none of which showed significant projections on the Numerical, Verbal and Spatial Factors. The USES (1970) concluded that Factor G represented a general reasoning ability since "it closely resembled the general reasoning factor found in studies conducted by the Army Air Forces in World War II". The USES eventually concluded that "intelligence" was an appropriate name for Factor G.

Upon the identification of the factors, the USES proceeded to select tests that would provide separate measures for each one of the ten factors. Tests were selected on the basis of two criteria: (1) internal or factorial validity; and (2) external or predictive validity. By selecting tests that met the orthogonal simple structure criteria and which predicted success in several occupations, the USES

eventually designed a battery of twelve tests that measured nine independent aptitudes. In a factor analytic sense, the nine aptitudes identified by the USES represent nine independent factors (the USES, 1970). The nine aptitudes measured by the 12 tests of the battery are: (G) Intelligence; (V) Verbal Aptitude; (N) Numerical Aptitude; (S) Spatial Aptitude; (P) Form Perception; (Q) Clerical Perception; (K) Motor Coordination; (F) Finger Dexterity; and (M) Manual Dexterity. The battery consists of eight pencil and paper tests and four apparatus board tests. Appendix A contains a description of the twelve tests and the nine aptitudes.

It is useful to examine the factor loadings of the twelve tests of the battery as some of the aptitudes are measured by more than one test. Unfortunately, the matrix of correlations from which the nine aptitudes were extracted was never published by the USES (1970). Table 1 presents the factor loadings of the twelve tests that were reported by the USES (1970). The Arithmetic Reasoning test shows relatively high loadings on Factors G and N. The tests measuring Three Dimensional Space and Vocabulary also show relatively high loadings on Factor G. The tests measuring Finger Dexterity and Manual Dexterity show high loadings on their respective factors. The tests measuring Form Perception and Three Dimensional Space appear to remain independent of one another. The Name Comparison test shows a high loading on the Clerical Perception Factor.

Table 1. GATB Factor Loadings

Aptitude Factor	Test	Factor Loading
G Intelligence	(3) Three Dimensional Space	.450
	(6) Arithmetic Reasoning	.552
	(4) Vocabulary	.513
V Verbal	(4) Vocabulary	.533
N Numerical	(2) Computation	.483
	(6) Arithmetic Reasoning	.438
S Spatial	(3) Three Dimensional Space	.500
P Form Perception	(5) Tool Matching	.520
	(7) Form Matching	.435
Q Clerical	(1) Name Comparison	.627
K Motor Coordination	(8) Mark Making	.423
F Finger Dexterity	(11) Assemble	.595
	(12) Disassemble	.486
M Manual Dexterity	(9) Place	.628
	(10) Turn	.500

Aptitude Factor Loadings reported in the USES (1970) Manual for the General Aptitude Test Battery, Section III: Development. Washington D. C: The United States Department of Labor.

In 1952, the USES developed GATB norms for the general working population. Normalized aptitude scales have a mean of 100 and a standard deviation of 20.

The score for Intelligence (G) was obtained by converting the raw scores yielded by Test 3 (Three-Dimensional Space), Test 4 (Vocabulary), and Test 6 (Arithmetic Reasoning) to normalized aptitude scores, and adding the three aptitude scores together.

The Numerical Aptitude (N) score was obtained by converting the raw scores yielded by Test 2 (Computation) and Test 6 (Arithmetic Reasoning) to aptitude scores, and adding the aptitude scores together.

The Form Perception Aptitude (P) score was obtained by converting the scores yielded by Test 5 (Tool Matching) and Test 7 (Form Matching), and adding the aptitude scores together.

The Finger Dexterity Aptitude (F) score was obtained by converting the raw scores yielded by Test 11 (Assemble) and Test 12 (Disassemble) to aptitude scores and adding the aptitude scores together. In a similar manner, Manual Dexterity (M) was computed with raw scores yielded by Test 9 (Place) and Test 10 (Turn).

The Verbal Aptitude (V) score was obtained by converting the total raw score yielded by Test 4 (Vocabulary) to an Aptitude score. The Spatial Aptitude (S) score was obtained in a similar manner with Test 3 (Three-Dimensional Space). The Clerical Perception (Q) scores was obtained by converting the total raw score yielded by Test 1 (Name Comparison) to an Aptitude score. Similarly, the Motor Coordination (K) score was obtained by converting the raw score yielded by Test 8 (Mark Making).

The norming sample consisted of 1,834 males, and 2,166 females. The average age of subjects was thirty years. The average level of education was completion of grade eleven. A majority of the sample lived in cities. Subjects worked in five occupational families which were: professional and semi-professional, office clerical and sales, skilled workers, equipment operators, and labourers. Data obtained from this population sample were used to construct a correlation matrix for the twelve tests of the battery (Table 2). The USES (1970) noted that tests measuring the same aptitude have high intercorrelations. The correlations ranged from .09 (between Vocabulary (Test 4) and Place (Test 9);

and Arithmetic Reasoning (Test 6) and Turn (Tests 10) to .78 (between Computation (Test 2), and Arithmetic Reasoning (Test 6)). Another high correlation is between Vocabulary (Test 4) and Arithmetic Reasoning (Test 6). The Vocabulary test also has high correlations with Name Comparison (Test 1) and Computation (Test 2).

Table 2. GATB Correlation Matrix

Test	1	2	3	4	5	6	7	8	9	10	11
1 Name Comparison											
2 Computation	.55										
3 Three Dimensional Space	.38	.45									
4 Vocabulary	.65	.64	.51								
5 Tool Matching	.64	.49	.47	.41							
6 Arithmetic Reasoning	.58	.78	.57	.72	.42						
7 Form Matching	.56	.48	.56	.45	.58	.46					
8 Mark Making	.59	.45	.24	.38	.45	.34	.42				
9 Place	.16	.15	.19	.09	.20	.14	.24	.33			
10 Turn	.24	.13	.16	.12	.25	.09	.27	.46	.52		
11 Assemble	.25	.17	.23	.15	.28	.15	.27	.33	.34	.38	
12 Disassemble	.32	.23	.22	.13	.35	.16	.33	.41	.44	.43	.47

Correlations obtained from the USES (1970) Manual for the General Aptitude Test Battery, Section III: Development. Washington, D. C: United States Department of Labor.

In 1983, the USES commissioned a study of the factorial structure of the GATB. Although the purpose of the investigation was to resolve differences arising from opposing models of the relationship between the aptitudes measured by the GATB and measures of performance on the job, the results of the study indicated that the nine aptitudes measured by the GATB break into three clusters

which the USES described as three general factors. The three general factors are a Cognitive Factor, a Perceptual Factor, and a Psycho-Motor Factor.

A confirmatory factor analysis of the correlations among the test scores of 23,428 workers was done by oblique multiple groups factor analysis. The USES (1983) reported that the factor structure underlying the GATB consists of three general factors: a Cognitive Factor (G, V, N,); a Perceptual Factor (S, P, Q); and a Psycho-Motor Factor (K, F, M).

Table 3 presents the USES (1983) correlation matrix from which the three general factors were extracted. The highest correlations are between Intelligence and Numerical (.86) and between Intelligence and Verbal (.84). The lowest correlation (.10) is between Manual Dexterity and Verbal. With a coefficient of .66, the correlation between Spatial Aptitude and Clerical Perception is high. The range of the correlations (.26 to .52) shared by Motor Coordination, Finger Dexterity, and Manual Dexterity is low to moderate. The correlations between Clerical Perception and: Intelligence (.64), Verbal (.62), Numerical (.66), and Form Perception (.65) are relatively high. The correlation between Spatial and Intelligence (.74) is high, but the correlations between Spatial and: Numerical (.51) and Verbal (.46) are moderate.

Table 3. GATB Aptitude Intercorrelations

Aptitude		G	V	N	S	P	Q	K	F
Intelligence	G								
Verbal	V	.84							
Numerical	N	.86	.67						
Spatial	S	.74	.46	.51					
Form Perception	P	.61	.47	.58	.59				
Clerical Perception	Q	.64	.62	.66	.39	.65			
Motor Coordination	K	.36	.37	.41	.20	.45	.51		
Finger Dexterity	F	.25	.17	.24	.29	.42	.32	.37	
Manual Dexterity	M	.19	.10	.21	.21	.37	.26	.46	.52

The table was obtained from the USES (1983) Test Research Report No. 44. The Dimensionality of the General Aptitude Test Battery (GATB) and the Dominance of the General Factors over Specific Factors in the Prediction of Job Performance for the U.S. Employment Service. Washington D. C: U.S. Department of Labor.

Table 4 presents the results of the USES (1983) confirmatory factor analysis of the aptitudes measured by the GATB. The USES noted that aptitude G was left out of the analysis since it was not defined independently of aptitudes V, N and S. The factor analysis shows that Spatial aptitude is farther from the Psycho-Motor aptitude than are the other two Perceptual aptitudes. Clerical Perception is closer to the Perceptual and Cognitive Factors than the Psycho-Motor Factor. Motor Coordination is closer to the Psycho-Motor and Perceptual Factors than the Cognitive Factor.

Table 4. Correlations between Aptitudes and General Factors

		General Factors		
Aptitude	Symbol	Cognitive	Perceptual	Psycho-Motor
Verbal	V	.82	.68	.32
Numerical	N	.82	.77	.42
Spatial	S	.59	.61	.35
Form Perception	P	.64	.81	.66
Clerical Perception	Q	.78	.81	.54
Motor Coordination	K	.48	.60	.64
Finger Dexterity	F	.25	.46	.67
Manual Dexterity	M	.19	.45	.72

The table was obtained from the USES (1983) Test Research Report No. 44. The Dimensionality of the General Aptitude Test Battery (GATB) and the Dominance of the General Factors over Specific Factors in the Prediction of Job Performance for the U.S. Employment Service. Washington D. C: U.S. Department of Labor.

In comparison with pervious studies, the USES (1983) noted that GATB studies which found three independent factors usually included three additional variables, namely, age, education, and socio-economic status. The USES (1983) concluded that although most orthogonal and oblique solutions appeared to be somewhat similar, orthogonal solutions often failed to show that Clerical Perception had a heavier loading on the Perceptual Factor than on the Cognitive Factor. Factor analysis indicated that the Perceptual Factor was almost totally predicted by the other two Factors, hence, aptitudes S, P, and Q were deemed to contribute very little to the predictive power of the battery. The USES reported that the Psycho-Motor aptitudes showed the lowest inter-aptitude correlations and the lowest aptitude reliability.

The USES (1970) reported that most of the reliability studies conducted with the GATB have been directed at the stability of measurement over time and at the equivalency of forms. The results published by the USES (1970) show the means and standard deviations of aptitude scores on initial testing and retesting as well as the correlations between them. In 1965-66, the USES conducted a test-retest reliability study of the aptitudes measured by the GATB with an interval of six weeks between initial testing and retesting. Sixteen State agencies participated in the study. Initial testing was done with GATB, B-1002, Form A, and retesting was done with an alternate form GATB, B-1002, Form B.

The sample consisted of 156 males and 168 females. The average age of males was approximately 32.7 years, with a standard deviation of 7.7 years. For males, the average number of years of education was approximately 11.8, with a standard deviation of 2.7 years. The average age of females was approximately 32.1 years, with a standard deviation of 9.6 years. For females, the average number of years of education was 12.4 years , with a standard deviation of 1.8 years. Table 5 presents the means, standard deviations, and reliability estimates for the GATB.

The GATB test-retest reliability estimates fall within a range of moderate (.67) to high (.95). For males and females, Intelligence (G), Numerical (N), and Verbal (V) show high stability over the six week interval between initial testing and retesting. The stability of Motor Coordination is identical for males and females, but Finger Dexterity and Manual Dexterity show a slight differentiation between males and females. Overall, the scores obtained by males and females on the second administration tend to be higher than the scores reported for the first administration. Over a six week period, the male scores tend to be slightly more stable than the female scores.

Table 5. GATB Test-Retest Reliability Estimates

		First Testing		Second Testing		Reliability
Aptitude		B-1001, Form A		B-1001, Form B		Six weeks
		M	SD	M	SD	r
G Intelligence	Males	105.3	19.1	105.6	21.2	.93
	Females	103.3	17.5	103.9	20.4	.90
V Verbal	Males	103.0	17.3	103.6	18.3	.90
	Females	106.4	17.5	107.1	19.0	.87
N Numerical	Males	103.6	21.6	103.9	24.8	.95
	Females	102.7	17.6	104.4	20.6	.90
S Spatial	Males	103.6	21.1	107.6	102.9	.88
	Females	98.6	18.1	102.9	18.3	.78
P Form Perception	Males	104.2	20.3	108.4	24.0	.84
	Females	105.3	21.0	112.3	21.0	.80
Q Clerical Perception	Males	110.7	20.1	116.9	22.7	.87
	Females	116.1	19.3	126.9	21.7	.84
K Motor Coordination	Males	101.2	21.6	109.7	24.0	.88
	Females	111.6	18.5	120.4	19.3	.88
Finger Dexterity	Males	94.1	22.0	97.8	21.4	.73
	Females	97.8	21.4	110.7	19.3	.67
Manual Dexterity	Males	98.6	24.5	106.2	19.9	.84
	Females	100.5	18.5	118.8	19.8	.72

Data obtained from the USES (1970) Manual for the General Aptitude Test Battery. Washington, D. C: The U.S. Department of Labor.

In 1986, the U.S. Department of Labor placed a two year moratorium on the GATB and commissioned the United States National Research Council to study the GATB (Baydoun and Neuman, 1992).

In terms of psychometric quality, the National Research Council (1989) noted that the lack of alternate forms and the speediness of tests make the GATB susceptible to score inflation due to coaching and guessing. The Council (1989)

suggested that "time limits not operationally linked to aptitudes should be altered as most examinees are able to complete only half of any one of the twelve tests in the time allowed". Baydoun and Neuman (1992) have suggested that "although the instrument could benefit from additional research, more comprehensible norms, additional versions, and more comprehensive score reporting techniques, the overall usefulness of the battery should not be in question".

In summary, the GATB consists of twelve tests which measure nine aptitudes. A matrix of correlations indicates that Intelligence (G) has high correlations with Verbal (V), Numerical (N), and Spatial (S) Aptitudes. Form Perception (P) has a moderately high correlation with Intelligence(G), Numerical (N), and Spatial (S) Aptitudes. Clerical Perception (Q) has a moderately high correlation with Intelligence (G), Verbal (V), Numerical (N), and Form Perception (P). Motor Coordination (K) has a moderately low correlation with all of the other aptitudes. Manual (M) and Finger (F) Dexterity have a moderately low correlation with each other and a low correlation with all of the other aptitudes. The confirmatory factor analysis conducted by the USES (1983) indicated that the nine aptitudes cluster into three general factors: Cognitive, Perceptual, and Psycho-Motor. Verbal and Numerical aptitudes show a high loading on the Cognitive Factor. Form Perception and Clerical Perception show a high loading on the Perceptual Factor, but Clerical Perception appears to be split between the Perceptual and Cognitive

Factors. Finger Dexterity and Manual Dexterity show high loadings on the Psycho-Motor Factor. The USES (1983) reported that although orthogonal and oblique solutions appeared to be somewhat similar, orthogonal solutions often failed to show that Clerical Perception has a heavier loading on the Perceptual Factor than on the Cognitive Factor.

Over a six week interval, the test retest reliability of the GATB was high, however, tests measuring Finger and Manual Dexterity tended to have lower test retest reliability than tests measuring the other ten aptitudes.

Use of the GATB with Hearing Impaired Students

Levine's (1974) survey identified the frequency with which some 91 tests were used in the psychological evaluation of deaf and hard-of-hearing children and adults. Approximately 99% of the respondents routinely conducted intelligence testing. Achievement testing accounted for 64% of the psychological testing. Personality testing was routinely conducted by 40% of the respondents. The procedures and format of testing included individual and group administrations, pencil-and-paper tests with multiple choice items, and non-language performance tests. Testing covered the age range from infancy through adulthood. Levine noted that a typical battery of tests included the Wechsler Performance Scales, the Stanford Achievement Test, and the Bender-Gestalt. Although the GATB was 33rd in Levine's rank-order of the 49 tests that were used by more-than-one respondent, the GATB was the only occupational aptitude test on the list. As the

GATB was only reported by five residential schools and one rehabilitation setting, the reported use of the GATB may have been related to studies sponsored by the United States Employment Service.

Traxler's (1989) survey of the frequency with which vocational tests were used for assessment in vocational training settings identified the GATB as being used to test 11% of the deaf students in academic and vocational courses. Traxler reported that the GATB was being used in many settings to measure the motor coordination, general ability, and intelligence of hearing impaired students. Over seven hundred studies of the GATB have been published since its development, but only a few studies involve administrations to deaf and hard-of-hearing persons (Berger, Holdt and Laforge, 1972; Botner, Stuckless, Moores, 1964; Botterbusch and Droege, 1972; Kronenberg and Blake, 1966; Moores, Harlow and Fisher, 1974; Sanderson, 1974). In each of the studies, the GATB was administered by several methods of communication. The methods of non-verbal communication used most often were gestures, demonstrations, pantomime, and signing systems that were never identified. Although non-verbal methods of communication allow researchers to obtain the test scores of deaf students, reliability estimates and the factor loading of measures obtained under modified conditions have never been reported in the literature.

Of the few available studies, the GATB data described by Botterbusch and Droege (1972) provide "one of the best sets of normative data" (Holm, 1987).

Between 1958 and 1969, the GATB was administered to deaf students in five residency schools. The total sample consisted of approximately 408 students. The size of the sample for each study ranged from 44 to 128 students. The age of the students ranged from 16 years to 22 years. Education ranged from 9 years of formal schooling to 11.5 years. The distribution of gender across the studies fluctuated from a minimum of 35% males to a maximum of 59% males. All of the students were administered a screening test, Non-Reading Aptitude Test Battery, and the GATB.

In all five of the studies, test administrators read the general instructions aloud while a school staff member interpreted the instructions in an unspecified sign language. Instructions to individuals were provided by hand written notes. Once the general instructions were given in the usual manner, the different schools used various methods to communicate instructions for each one of the tests. On two occasions, directions were written on a blackboard or presented through the use of prepared charts. One school provided the students with a written description of the signals that would be used during testing. In one instance, the test administrator "used slow demonstration, with flourishes to emphasize certain points" while another person read the directions aloud and a school staff member provided simultaneous interpretation. The results of the study for the total sample are shown in Table 6.

According to Botterbusch and Droege (1972), students performed well above average on aptitudes S, P, and Q (perception); and close to average on aptitudes K, F, and M (performance). However, scores on aptitudes G, V, and N (cognition) ranged from 1 to 1.5 standard deviations below the general working population mean of 100. The most likely explanation for low scores on Verbal and Numerical tests was that two of the tests required an English language reading grade level well above the reading grade level of most deaf students. The 6.0 reading grade level of most items (USES, 1970) may have been too difficult, "as the average deaf student at age 17 is only reading at grade level 4.0" (Morgan and Vernon, 1994).

Table 6. GATB Aptitude Means and Standard Deviations

GATB Aptitude	Combined Total of the Five Studies Means	Standard Deviation
G General Learning	80.1	17.0
V Verbal	73.4	7.3
N Numerical	84.9	21.5
S Spatial	104.5	23.6
P Form Perception	112.4	22.5
Q Clerical Perception	103.7	22.2
K Motor coordination	104.9	23.2
F Finger Dexterity	101.9	29.1
M Manual Dexterity	105.2	22.7

The data presented in Table 6 was obtained from Botterbusch and Droege (1972). GATB aptitude testing of the deaf: problems and possibilities. Journal of Employment Counselling. vol. 9. no.1, (14-19).

Botterbusch and Droege (1972) suggested that these results were a "strong indication that modifications in administration procedures to meet the needs of the deaf were successful ...[and]... that the deafness of the test subjects did not limit their performance on the entire GATB." Although Botterbusch and Droege described the distribution of the scores on aptitudes G, V, and N, they did not interpret the scores. Sanderson (1974) noted that the results of the five state study were typical of the results of other administrations of the GATB to hearing impaired students.

Sanderson's (1974) study of Motor-Coordination (K), Finger Dexterity (F), and Manual Dexterity (M) of students who had been educated orally and manually showed no significant differences in the mean score of the two groups. In administering tests orally, test administrators read directions loudly and slowly. Once directions had been read aloud, examinees were told to read the printed directions. Following the reading of the directions, tests were demonstrated and examinees completed practice exercises. Testing commenced when examinees demonstrated mastery of the test taking tasks. With manually oriented students, tests were administered by signing or finger spelling. Pantomime was used to instruct students who were unable to read lips or understand signs. However, Sanderson only used pantomime to instruct students to read the printed directions. Sanderson argued that flexibility in administration would not affect the scores of students who demonstrated mastery of the test taking tasks.

Sanderson reported that the use or non-use of manual communication with deaf students did not affect the measurement of aptitudes K, F, and M. The scores of the students on these aptitude scales were equivalent to the mean of the general working population. Sanderson observed a sex difference in Finger Dexterity (F). However, as the difference was not attributed to the method of communication, it was ignored. Sanderson suggested that the difference was simply a matter of females being better at Finger Dexterity than males.

Sanderson also suggested that Verbal and Numerical scores indicated the standing of deaf students relative to the mean of the general working population. In so far as the English language reading grade level of Verbal and Numerical items are above the reading level of most deaf adults, the validity of Sanderson's opinion is questionable.

Canadian Edition of the GATB

In 1971, Employment and Immigration Canada published GATB scores for the occupations listed in the Canadian Classification and Dictionary of Occupations (CCDO). In the introduction to the CCDO the editors noted that "the task of relating an applicants' qualifications to occupational requirements, or counselling [the individual] in terms of the greatest likelihood of job success is facilitated through the use of test norms and critical scores.... Actual Canadian test norms, however, are not yet available except for a relatively small number of the thousands of occupations in the Canadian economy" (CEIC, 1971). To

compensate for the lack of empirical data, the editors of the CCDO reported that as a result of a series of comparative studies, "the aptitude ratings found in the qualification profile of volume two of the CCDO reflects judgments, based on available data, as to the average level of aptitude required for satisfactory performance in an occupation " (CEIC, 1971).

With the introduction of the metric system to Canada, Employment and Immigration Canada deleted all references to units of Imperial and Metric measurement from GATB, Form A (CEIC, 1986). In 1986, the revised Form A was used to develop norms for the general working population of Canada. In the 1986 Canadian edition of the manual for administering and scoring the GATB, Factor G is described as General Learning Ability rather than Intelligence. Although Canadian norms for the GATB have a mean of 100 and a standard deviation of 20, a description of the norming sample is unavailable.

Summary

In summary, the difficulties encountered in the administration of standardized tests to deaf persons have been attributed to two main factors. One of the factors involves the format and procedures of pencil and paper tests with multiple choice items. The other factor involves the reading grade level of test items.

The effect of familiarity with a method of communication on performance is that deaf children who are familiar with both the method of communication and the format and procedures of testing tend to experience little difficulty performing

test taking tasks in the prescribed manner. Bragman's (1882a) study indicates that pantomime, demonstration, and simultaneous communication may not be interchangeable methods of communicating instructions to deaf children. Morgan and Vernon (1994) have asserted that American Sign Language (ASL) is often superior to pantomime and demonstration in the conveying of test instructions to deaf children. Although ASL has been identified as a superior method of communicating test directions to deaf adults, ASL and Siglish (Signed English) have been shown to be interchangeable methods of conveying test directions.

One of the troublesome aspects of pencil and paper tests with multiple choice items is that many deaf adults experience difficulty matching the spaces printed on separate answer sheets with the items printed in test booklets. Further, administrative difficulties tend to become exacerbated by time limits. In addition, the format and procedures of a paper and pencil test with multiple choice items have not been shown to be equivalent to the format and procedures of a videotaped presentation of identical items conveyed in ASL.

The difficulty with the reading grade level of items is that most deaf adults read at a grade level well below the reading grade level that is required by many standardized tests. In other words, Intelligence tests which involve the use of the English language often measure the language deprivation of deaf examinees.

In terms of construct validity, the correlation between tests that measure the same aptitude tends to be high, and low for tests that measure dissimilar aptitudes. The twelve tests of the GATB load onto nine independent factors

which cluster into three general factors. In terms of factor structure, the USES (1983) noted that solutions obtained by oblique and orthogonal rotations tended to be very similar.

With the exception of Manual Dexterity and Finger Dexterity, test-retest reliability estimates were high. Also, an increase in mean scores was observed for a six week interval between initial testing and retesting.

The National Research Council (1989) noted that the lack of alternate forms and the speediness of tests make the GATB vulnerable to coaching and guessing. Further, the time limits established for the GATB may be too severe as most examinees complete only half of the items of any one of the tests in the time allowed.

Administrations of the GATB by Botterbusch and Droege (1972) to deaf students indicated that the measures of Verbal and Numerical Aptitude were approximately 1 to 1.5 standard deviations below the general working population mean of 100. The performance of deaf students on the other measures fell within 1 standard deviation of the general working population mean of 100. However, the reliability and the validity of the measures yielded by administrations of the GATB to deaf persons have never been reported in the literature.

Chapter 3

Methods and Procedures

Methods and procedures are described in three sections: the development of the GATB-ASL, data collection, and data analysis. The development of the GATB-ASL includes descriptions of the criteria for translating the directions into ASL, the translation procedure, the translation review procedure, the video production procedure, a description of the GATB-ASL, and administration procedures. The section on data collection includes a description of the sample, and the scoring procedures. The section on data analysis identifies the methods that were used to estimate reliability and to conduct the exploratory factor analysis.

Development of the GATB-ASL

Translation Criteria

With respect to communicating instructions, the Standards for Educational and Psychological Testing indicate that instructions should be "presented with sufficient clarity and emphasis so that it is possible to approximate for others the administrative conditions under which the norms and the data on reliability and validity were obtained" (AERA, APA, NCME, 1985). Further, the standards indicate that the "directions presented to test takers should be detailed enough so that test takers can respond to a task in the manner that the test developer intends" (AERA, APA, NCME, 1985). The basic rationale of these standards is

that the clarity, emphasis, and detail of instructions should be such that everyone performs the same test taking tasks in the same manner.

Empirical evidence in support of the argument that the directions conveyed in ASL have the same clarity, emphasis, and detail as the directions conveyed in the English language came from two sources. One source of evidence was a comparison of the instructions given in ASL with the instructions given in the English language. A comparison of a translation with an original text provides evidence for the accuracy, clarity, and emphasis of the translation, however, such a comparison does not show that test takers have been provided with enough detail so as to be able to replicate the intentions of test developers. To show that instructions conveyed sufficient detail, the performance of test takers was observed during administrations of the GATB-ASL.

Translation Procedure

Prior to translating the instructions into ASL, a panel of representatives of deaf culture stakeholder groups were asked to recommend ASL signs that conveyed the intentions of the developers of the GATB. An example may help to clarify the recommendation process. In the phrase "Ready, begin", the word "begin" was signed as "go ahead" rather than "start". The sign for "go ahead", was selected because it denotes the idea of proceeding; whereas, the sign for "start" denotes the starting of a car.

All of the signs that were used in the ASL translation were described in the *Perigee Visual Dictionary of Signing* (Butterworth, 1983). As the Perigee Visual

Dictionary usually provides more than one English language word for any given sign, every effort was made to separate semantic differences from substantive translation errors. A substantive translation error was deemed to consist of either an inaccurate direction or an inappropriate use of sign. The appropriateness of signing was determined by the conformity of signing with the language structures presented in the American Sign Language Series: *Signing Naturally* (VISTA, 1988).

Two attempts at presenting verbal instructions in ASL were made before the final version of the GATB-ASL was recorded on videotape. The first attempt consisted of modifying ASL to conform with the grammatical structure of the English language. The transliteration approach consisted of substituting signs for words. The transliteration approach could not easily be done because of the temporal and spatial differences between the grammatical structures of ASL and English. The second attempt consisted of signing only the main ideas in the English language text. The main ideas of the English language text were reduced to key words which were shown on a tele-prompter. The test administrator was supposed to provide an ASL interpretation of the key-words as they appeared on the screen of the tele-prompter. This approach was similar to the procedures that were used to provide an ASL interpretation of the House of Commons' Question Period. As the display of the key words never matched the on-camera signing, the second attempt was a stiff and awkward. The final attempt involved training a deaf person to administer the GATB, and asking her

to modify the administrative procedures to accommodate deaf test takers. The training of the test administrator and the modifying of the procedures were accomplished by a test development team.

The team consisted of a team leader, a deaf test administrator, a deaf test subject, an ASL interpreter, a certified GATB test administrator, and a deaf video camera operator. The team decomposed the GATB instructions into 90 ASL segments of which 24 repeat the imperative " do not write in the test booklet; do not turn the page until you are told to do so" (CEIC, 1986). When the test administrator demonstrated mastery of the ability to administer a specific segment to the deaf test subject, the signing of that segment was recorded on video.

The following text represents a typical example of the translations produced by the team. The original English language text for Test 8, Mark Making, reads as follows:

> Stop: Do not turn the page until told to do so.
> Now look at the instructions at the bottom of the page while I read them aloud. On the back of this page is another exercise exactly like this. Work as fast as you can. This time you will have 60 seconds.
> Now turn the page and wait for the signal to begin.
> Hold your pencil ready for marking.

An English language rendition of the team's ASL translation reads as follows:

> First practice finish; second real test; turn page; same before
> First, second mark same, left to right
> OK; now real test; try work faster
> Start point (points to page); mark lines; time limit 60 seconds
> Knows Counsellor (points to the side) inform go ahead

The team verified the accuracy of the translation by composing an English language translation of the ASL text. The team's reverse translation of the Mark Making instructions conveyed in ASL is presented in the following text:

> Now we've completed the first and second practice segments. Now turn over the page for the actual test. Follow the same process as you did for the practices one and two. You will start in the first square, work your way across the rows from left to right. try and work even more quickly this time. You will be allowed sixty seconds. the test administrator will tell you when to begin.

Translation Review Procedure

The clarity, and emphasis with which the English language text was conveyed in ASL was determined by two methods. One of the methods involved a comparison of a back-translation of the ASL text with the original English language text. The other method involved a direct comparison of the ASL text with the English language text.

Upon completion, the GATB-ASL was given to three ASL interpreters who had no prior involvement with the activities of the test development team. The interpreters were asked to produce an English language translation of the ASL presentation.

For the Mark Making example, the back translation resulted in the following text:

> We have completed the first and second segments. Now we'll go on to the actual test. Turn the page over. Follow the same process as you did for the practice exercises one and two. You will start in the first square, marking the three lines in each square and working across the rows from left to right. During this test, try to work even more quickly. You will be allowed sixty seconds. The test administrator will tell you when to begin.

A panel composed of individuals who were not involved with the development of the GATB-ASL compared the ASL translation, the back translation, and the original text. Panel members compared the translations and submitted their judgments to the test development team leader. The test development team usually complied with the review panel's suggestions for revisions.

The adequacy of the directions conveyed in ASL was determined by observing the response of deaf test takers to the directions. For example, if a GATB direction required test takers to "blacken the space under the correct letter in the Part 2 Practice section" (CEIC, 1986), and the response to the direction given in ASL was that the deaf test takers picked-up their pencils and blackened the correct space in the correct section of the answer sheet; the observed response was deemed to be congruent with the directions. The perceived congruence was accepted as evidence that ASL conveyed the direction with sufficient detail so as to approximate the administrative condition intended by the developers of the GATB.

Video Production Procedure

A video production team created computer animated graphics to display the items and the answers of the practice exercises for tests 1 to 7. Also, graphics were used to give on-screen instructions to test administrators. The on-screen graphics usually contained the imperative to stop the tape.

In the final production phase, a verbal narration was edited to the video. As speaking time did not match signing time, it was necessary to modify the English language text. With respect to Test 8, Mark Making, the English language text was modified to the following:

> We've completed the first and second practice. Now do the actual test. Turn the page over. Follow the same process used for practice exercises one and two. Start here, mark three lines, work across the rows from left to right. This time try to work even more quickly. You will be allowed sixty seconds. The Counsellor will tell you when to begin.

Signing and speaking were synchronized so that speaking slightly followed signing. In fact, every effort was made to prevent speaking from preceding signing. While a voice narration was not essential, it was decided that it may be useful to some hard-of-hearing persons. Another reason for the voice narration was the spatial convention that was used in signing. The spatial convention requires test administrators to stand to the right of the TV and face the test takers. If this convention is not followed some deaf persons may experience confusion with the instruction to look at the Counsellor to the right of the TV. Given that most of the counsellors with deaf clients tend to be hearing, it was thought that it would be fairly easy for them to stand next to the TV and to attend to the instructions by listening to the voice narration.

An explanation for using the sign for Counsellor, in lieu of the sign for test administrator, may be appropriate. Throughout the development of the GATB-ASL, the team used the sign "test + person" for test administrator. Unfortunately, the title test administrator was too long for the graphics and for the amount of

time that was allowed to speak the phrase: "the test administrator will tell you when to begin". Eventually, the sign for "Counsellor" fell into place as an appropriate substitute for the term "test administrator".

Silence is a poignant aspect of the video presentation. The graphics and a few of the ASL segments are presented in silence. To the uninitiated, this may seem somewhat unusual, but eventually one becomes accustomed to the silence. One aspect of the video presentation in which voice was selected over silence involves the demonstrations of the manipulation of the pegs and the rivets.

A master tape was made by combining 130 computer generated graphics with 90 ASL segments and 15 demonstrations. Copies of the master tape were reproduced on a standard VHS tape.

The GATB-ASL

The American Sign Language version of the General Aptitude Test Battery (GATB-ASL) consists of 12 tests. Tests 1 to 8 are pencil and paper tests, and tests 9 to 12 involve the use of apparatus boards. The items of tests 1 to 7 are presented in test booklets. The number of items per test vary from 25 items to 150 items, time limits vary from five minutes to seven minutes. To take these tests, examinees look at the items in the test booklets, select their answer from among the multiple-choices, and mark their answers on a separate answer sheet. Test 8 is making the same mark [Ⅱ] as many time as possible in one minute. Test 9 involves placing 48 cylindrical pegs in holes. Test 10 is turning-over the

pegs. Tests 11 is putting rivets and washers together. Test 12 is taking them apart. Time limits for tests 9 to 12 fall within a range of 45 seconds to 90 seconds. Before taking each test, examinees attend to instructions, observe demonstrations, and complete a practice exercise.

GATB-ASL Administration Procedures

The procedures consist of conveying instructions in ASL, using animation and graphics to show the correct responses to the practice exercises, and using ASL to describe the demonstrations immediately prior to showing them. All of the instructions, animation, graphics, descriptions, and demonstrations were recorded on a standard VHS videotape. Procedures for testing right handed and left handed people were recorded on separate tapes. To administer the battery, test administrators start and stop the videotape; and start and stop testing. Test items, test booklets, answer sheets, peg boards, rivet boards, time limits and the order of test administration are identical to the GATB.

With the exception of the TV, a room was set-up according to the instructions presented in the manual for administering and scoring the GATB. The TV was placed in the front-centre of a room so that test takers had an unobstructed view of the screen and such that test administrators could check a person's answers without obstructing another person's view of the TV. The battery of 12 tests was administered to groups of four to six persons by one test administrator who was assisted by an ASL interpreter.

Test administrators commenced testing by signing "Please, look at the TV". The TV was turned-on and the videotape was played. When instructions for a practice exercise were completed, the tape was stopped. Test administrators commenced practice exercises by signing: "Practice exercise, do you have any questions, do you understand, ready, begin". At the end of a practice exercise, the correct responses were shown by starting the tape. While test takers compared their responses to the correct responses, each person's work was examined by the test administrator. To communicate with test subjects who had incorrect responses, test administrators used the services of the ASL Interpreter. Upon completion of a practice exercise, instructions for the test were shown. When the instructions were completed, the videotape was stopped. Testing was started by signing: "Do you have any questions, do you understand, ready, begin". Testing was stopped by flashing the ceiling lights. Tests 8 to 12 were both started and stopped by flashing the ceiling lights. Starting and stopping procedures continued until the 12 tests of the battery had been administered.

The time required to administer the battery was approximately 2.5 hours. Examinees received a five minute rest period every 45 minutes. Right and left handed persons were tested separately. In addition to obtaining test scores, test administrators were asked to record their observations. At the end of the second test session, subjects were asked questions about their education, employment, age at the on-set of deafness, and degree of impaired hearing. Subjects were

informed that disclosure of personal information was voluntary. Everyone was told that scores could not be used to match aptitudes with occupations but feedback on performance would be made available four weeks after testing.

As the methods of communicating directions to hearing people and the methods of communicating directions to deaf people are seldom the same, the conditions of verbal testing are hardly ever equivalent to the conditions of non-verbal testing. As a consequence, one cannot assume that the measurements obtained under non-verbal conditions were unaffected by the modifications that accommodated deafness. Thus, a critical issue was the effect of the modified procedures on the measurements being taken. That is, too what extent were the scores obtained under the non-verbal conditions deemed to be reliable and valid measures of the constructs measured by the GATB.

Data Collection

Data collection consisted of administering the GATB-ASL to the same test subjects on two occasions with an interval of six weeks between initial testing and retesting.

Sample

The sample consisted of 106 deaf and hard-of-hearing adults eighteen years of age or older who were fluent in ASL. The definitions of deaf and hard-of-hearing that were used in this study correspond to the categories of hearing impairment devised for the Health and Activities Limitations Survey (HALS) of

Canadians with Impaired Hearing (Schein, 1992). The term 'deaf' was understood to mean Category III which includes "those who are completely unable to hear in one-person conversations". The term 'hard-of-hearing' was understood to mean Category II which includes "those who say they have partial difficulty hearing one person and have at least partial difficulty hearing in groups". Individuals were not included in the sample if they used a technical aid which eliminated their hearing limitation. Individuals with more than one limitation were excluded from the data base (e.g., a very small number of deaf-blind adults were tested but their scores were not included in the final analysis).

The selection criteria were that deaf and hard-of-hearing adults were 16 years of age or older and were fluent in ASL. The selection criteria did not specify age at on-set of deafness, number of years of formal education, the number of persons by gender, nor employment status.

Agencies with deaf and hard-of-hearing clients and schools for the deaf were asked to recruit test subjects. Three agencies and one school responded to the request. They assisted with the task of recruiting test subjects and to a follow-up request for space to administer the test. For ease of description, the four organizations that assisted with the study shall simply be referred to as test centres. The test centres recruited deaf and hard-of-hearing persons by a variety of different methods including letters, posters, informal presentations at social events, and personal requests made by members of the staffs at the test centres.

The test centres recruited approximately 120 volunteers. Approximately thirty volunteers were recruited in each of four provinces which included Nova Scotia, Ontario, Alberta, and British Columbia.

ASL Interpreters at the test centres were asked to determine the fluency with which volunteers communicated in ASL. Each volunteer was screened by means of a personal interview which usually consisted of a discussion of the study and a preview of some of the items presented in the video.

To evaluate the screening process and the accuracy of the ASL translation, 10% of the test subjects were randomly assigned to a pre-test group. The entire battery of tests was administered to the pre-test group. The persons who administered the GATB to the pre-test group were not the same persons who screened the volunteers. The scores obtained during the pre-testing were not included in the data that were used to compute reliability and the factor analysis. Also, the test subjects who participated in the pre-testing were excluded from further participation in the study.

Scoring

The raw scores of each test reflect the total number of correct responses to the test items. The total number of correct responses for test 9 and 11 was determined by counting the empty holes in the upper part of the apparatus boards. The total number of correct responses for test 10 included pegs which had been turned over plus pegs held by test takers at the end of the test. The

number of correct responses for test 12 was determined by counting the empty holes in the lower part of the apparatus board.

All of the answer sheets were hand scored by one person who used the GATB, Form A, Scoring Template (Nelson Canada, 1986). A second person verified the accuracy of the hand scoring. Overall, the re-scoring of 88 randomly selected answer sheets from among the 240 showed that the raw scores had been accurately tabulated the first time. By using the score conversion tables in the Manual for the administration and Scoring of the GATB (CEIC, 1986), raw scores were converted to normalized aptitude scores with a mean of 100 and a standard deviation of 20.

Data Analysis

Data analysis included estimates of internal consistency, test-retest reliability, and an exploratory factor analysis.

Reliability Estimates

The test-retest design yielded two sets of data which corresponded to the responses of the same test subjects to the same items on two occasions. The interval between testing was six weeks. The Pearson Product Moment Correlation between the scores yielded by the two administrations provided an estimate of test-retest reliability. Pencil and paper tests with multiple choice items were dichotomously scored. Estimates of the internal consistency of the seven pencil and paper tests were calculated with the Kuder Richardson Formula

Number 20 (KR-20). The KR-20 was not calculated for the five performance tests as each result was the number of times single tasks were performed correctly. The standard error of measurement was also calculated.

Factor Analysis

An exploratory factor analysis was done with the SPSS for Windows 3.1 program. Principal component factors with eigenvalues greater than one were rotated orthogonally to a simple structure. Appendix C contains a copy of the factor analysis computer output.

Chapter 4

Results

The results have been presented in three sections which correspond to observations of the performance of the hearing impaired sample, descriptions of GATB-ASL scores, and an exploratory factor analysis. Before presenting the results, the sample is described.

Sample

Usable data were obtained from 106 of the 120 people who were tested. Approximately 80% of the test subjects were completely unable to hear in one-person conversations; their degree of impaired hearing was deemed to correspond to category III of the Health and Activities Limitation Survey (HALS) (Schein, 1992). Approximately 20% of the sample reported partial difficulty hearing one person and at least partial difficulty hearing in groups; their level of impaired hearing was deemed to correspond to category II of the HALS. Females accounted for approximately 52% of the sample and males for approximately 48% of the sample. The age of the sample ranged from 16 years to 63 years of age. The sample mean was 32 years of age, with a standard deviation of 11 years. The sample median age was 32 years of age. The sample mode was 18 years of age. Twenty-three subjects were 16 to 19 years of age. Twenty-two subjects were 20 to 29 years of age. Thirty-five subjects were 30 to 39 years of age, and twenty-three subjects were 40 years of age or older. The data on education

reflect each person's best estimate of the highest completed grade level of formal education. The grade levels of education which had been completed at the time of data collection (September, 1993) ranged from grade 6 to one year of graduate studies. The sample mean was grade 11, with a standard deviation of 2 grade levels. The sample median was grade 11. The mode was grade 12. Fourteen subjects had completed grade 9 or lower. Fifty-eight subjects had completed grades 10, 11 or 12. Twenty-three subjects continued beyond grade 12 to pursue either a college diploma or a university degree. The data on employment contained too many missing values to warrant analysis. Table 7 contains a description of the sample by the categories of impaired hearing, gender, age, and educational grade level.

Table 7. GATB-ASL Sample.

Category of Impaired Hearing		Gender		Age		Education	
Category II	Category III	Female	Male	Mean	S.D.	Mean	S.D.
21	85	56	50	32	11	11	2

N = 106

Performance of the Sample

The performance of the hearing impaired sample was observed during both administrations of the GATB-ASL. Overall, the participants demonstrated very little difficulty following the directions conveyed in ASL. The test taking task of matching the spaces printed on the answer sheet with the items printed in the test booklets was usually performed quickly and accurately.

A majority of the test subjects expressed annoyance at the continual repetition of the command: "Remember, do not write in the test booklet, do your work on the scrap paper, mark your answers on the answer sheet". During these instructions, test subjects would look away from the TV screen, converse with each other, fiddle with their pencils, tap the tables, move their chairs, and stand-up and stretch. The test administrator banged a table to regain the attention of the group and used rapport to settle the group prior to starting a test.

For the practice exercises of Tests 1 through 7, the test administrator compared each person's answers with the correct answers displayed on the TV screen. Many of the test subjects were guarded about showing their answers to the test administrator. It was not uncommon for participants to correct their addition, subtraction, and multiplication mistakes before showing their answer sheets to the test administrator. Participants did not hesitate to help each other with tasks, such as finding solutions to the arithmetic word problems in Test 6 or explaining the similarities and differences between the words listed in Tests 1, and 4. A few participants attempted to fill-in their friend's answer sheet. When interrupted, these helpful individuals would explain that giving answers to difficult problems was not only appropriate behavior but expected behavior among deaf friends. In some instances, there appeared to be no difference between helping a friend and doing a friend's work. It is important to note that neither the test administrators nor the test subjects considered these helpful behaviors to be

intentional acts of deception. These actions were simply regarded as overly helpful behaviors.

In response to these helpful behaviors, the test administrator would stop the practice exercise and explain that the purpose of testing was to determine the ability of each person to answer the questions without help. The administrator suggested that it was OK to skip questions that were too difficult, and to make a guess before going on to the next question. These suggestions were followed by questions about the penalty for skipping or guessing answers. On these occasions, the test administrator would explain that only correct answers were counted, hence, guessing and skipping were appropriate test taking strategies. A few persons followed the suggestions by filling-in the answer sheets for every item in Test 4 (Vocabulary) and Test 6 (Arithmetic Reasoning) without looking at the questions. These individuals did the practice exercise in the proper manner but ignored the test questions when filling-in the answer sheet during the actual test. When asked about their behavior, they usually explained that the vocabulary was too difficult or that they did not understand the arithmetic word problems. Only one person explained that she had not been taught to multiply or divide and that she felt too embarrassed to turn in an almost completely blank

answer sheet. Appendix B contains descriptions of the questions and the problems encountered in the administration of the GATB-ASL.

To a hearing person, groups of deaf test takers tend to be very noisy. During testing, tables were banged, chairs were squeaked, pencils were tapped, hands were slapped, and individuals made loud audible sounds, such as laughter. As noise does not disrupt testing, no effort was made to reduce the noise level.

GATB-ASL Scores

Raw Test Scores

Table 8 presents raw score means, standard deviations, and mean differences yielded by two administrations of the GATB-ASL with an interval of six weeks between initial testing and retesting.

The mean differences ranged from -1.43 (Test 4, Vocabulary) to 5.35 (Test 1, Name Comparison). The mean differences of Test 4 (Vocabulary), Test 6 (Arithmetic Reasoning), and Test 11 (Assemble) were not statistically significant. The mean differences of the remaining nine tests were statistically significant. Overall, GATB-ASL scores have been shown to have a tendency to increase over the six week interval between initial testing and retesting.

Table 8. GATB-ASL Means, Standard Deviations, and Mean Differences.

Tests of GATB-ASL		Time 1		Time 2		Mean
No.	Name of test	Mean	S. D.	Mean	S.D.	Difference
1	Name Comparison	48.90	15.87	54.25	18.83	05.35 **
2	Computation	17.64	06.46	18.56	06.94	00.92 **
3	Three-Dimensional Space	18.57	06.72	21.85	07.01	03.28 **
4	Vocabulary	12.23	06.95	10.80	06.28	-01.43 *
5	Tool Matching	30.52	07.77	33.16	08.37	02.64 **
6	Arithmetic Reasoning	06.95	03.67	07.33	03.83	00.38 *
7	Form Matching	27.48	09.38	30.33	10.56	02.85 **
8	Mark making	70.52	10.97	74.99	11.44	04.47 **
9	Place	89.29	10.86	93.33	11.63	04.04 **
10	Turn	94.90	13.89	101.39	13.49	06.49 **
11	Assemble	27.00	06.35	29.27	05.37	02.07 *
12	Disassemble	26.94	05.50	29.58	04.77	02.64 **

N = 106; * P> .05; ** P< .05

Reliability of Test Scores

Table 9 presents estimates of internal consistency, test-retest reliability, and standard error of measurement. Reliability estimates were computed with data obtained from 106 deaf and hard-of-hearing adults on two administrations of the GATB-ASL.

The internal consistency of the paper and pencil tests with multiple choice items (Tests 1 to 7) was estimated with Kuder-Richardson Formula Number 20 (KR-20). For the first administration (time 1), internal consistency ranged from .80 to .96; and standard error of measurement ranged from 1.61 to 2.69. For the second administration (time 2), internal consistency ranged from .80 to .97; and standard error of measurement ranged from 1.68 to 3.26. Internal consistency was not estimated for the performance testing (Tests 8 to 12) as these tests measure the ability to perform single tasks during very brief time trials.

Estimates of test-retest reliability reflect the correlation between responses to items from two administrations with an interval of six weeks between initial testing and retesting. Test-retest reliability estimates ranged from .26 to .91. Estimates for the pencil and paper tests with multiple choice items (Tests 1 to 7) ranged from .68 to .91; the standard error of measurement ranged from 1.61 to 5.93. For performance testing (Tests 8 to 12), estimates of test-retest reliability ranged from .26 to .84; the standard error of measurement ranged from 2.53 to 5.72.

Table 9. GATB-ASL Reliability Estimates.

Tests of GATB-ASL		Internal Consistency				Test-retest	
		Time 1		Time 2			
No.	Name of test	KR-20	SEM	KR-20	SEM	P_{tt}	SEM
1	Name Comparison	.96	2.69	.97	3.26	.86	5.93
2	Computation	.91	1.93	.91	2.08	.91	1.91
3	Three-Dimensional Space	.86	2.48	.88	2.38	.83	2.56
4	Vocabulary	.89	2.29	.89	2.07	.77	2.72
5	Tool Matching	.93	2.02	.93	2.17	.68	3.71
6	Arithmetic Reasoning	.80	1.61	.80	1.68	.74	1.61
7	Form Matching	.94	2.25	.95	2.32	.79	3.71
8	Mark Making					.82	4.25
9	Place					.74	4.83
10	Turn					.76	5.72
11	Assemble					.26	2.55
12	Disassemble					.46	2.53

The data in Table 9 was originally published in Cooney (1994). Manual for the Administration of the General Aptitude Test Battery in American Sign Language. Ottawa, Ontario: Algonquin College of Applied Arts and Technology.

Aptitude Scores

The Canadian Manual for Administering and Scoring of the General Aptitude Test Battery (CEIC, 1986) contains tables for converting raw scores to normalized aptitude scores. Normalized aptitude scores have a mean of 100 and a standard deviation of 20.

The General Learning Aptitude (G) score was calculated by converting the raw scores yielded by Test 3 (Three-Dimensional Space), Test 4 (Vocabulary), and Test 6 (Arithmetic Reasoning) to normalized aptitude scores, then adding the aptitude scores together.

The Numerical Aptitude (N) score was obtained by converting the raw scores yielded by Test 2 (Computation) and Test 6 (Arithmetic Reasoning) to aptitude scores, and adding the aptitude scores together.

The Form Perception Aptitude (P) score was obtained by converting raw scores to aptitude scores for Test 5 (Tool Matching) and Test 7 (Form Matching), and adding the aptitude scores together.

The Finger Dexterity Aptitude (F) score was obtained by converting raw scores to aptitude scores for Test 11 (Assemble) and Test 12 (Disassemble), and adding the aptitude scores together. In a similar manner, Manual Dexterity (M) was computed with raw scores yielded by Test 9 (Place) and Test 10 (Turn).

Verbal Aptitude (V) was measured by Test 4 (Vocabulary). Spatial Aptitude (S) was measured by Test 3 (Three-Dimensional Space). Clerical Perception (Q) was measured by Test 1 (Name Comparison); and Motor Coordination (K) was measured by Test 8 (Mark Making).

For the nine aptitudes measured by the GATB-ASL, estimates of test-retest reliability ranged from .40 to .92 with the standard error of measurements ranging

from 5.72 to 12.57. Table 10 presents the Aptitude means, standard deviations, mean differences, and test-retest reliability estimates.

The range of the scores for aptitudes G, V, and N were 1.5 to 2 standard deviations below the general working population mean of 100. All of the other aptitude scores fell within 1 standard deviation of the general working population mean.

Table 10. GATB-ASL Aptitude Means, Standard Deviation, Mean Difference, Reliability

Aptitude Scale	Time 1		Time 2		Mean	Test-retest	
Symbol Name	Mean	S. D.	Mean	S.D.	Difference	r_{tt}	SEm
(G) General Learning	69.49	16.44	74.53	16.28	5.04	.85	5.73
(V) Verbal	62.08	17.69	66.29	17.35	4.21	.77	7.22
(N) Numerical	72.93	22.16	76.12	24.10	3.19	.92	6.20
(S) Spatial	97.83	21.33	108.64	22.69	10.81	.83	8.21
(P) Form Perception	98.07	21.28	103.92	24.09	5.85	.80	8.91
(Q) Clerical Perception	90.16	19.71	96.81	23.42	6.65	.86	7.37
(K) Motor Coordination	102.06	21.29	110.68	22.19	8.62	.82	8.26
(F) Finger Dexterity	90.83	27.84	102.04	23.97	11.21	.40	12.57
(M) Manual Dexterity	95.05	26.04	107.02	26.78	11.97	.76	7.82

The data in Table 10 was originally published in Cooney (1994). Manual for the Administration of the General Aptitude Test Battery in American Sign Language. Ottawa, Ontario: Algonquin College of Applied Arts and Technology

Exploratory Factor Analysis

Table 11 presents a correlation matrix for the 12 tests of the GATB-ASL. An inspection of the matrix indicates that the correlation coefficients range from .06 to .75. The highest correlation (.75) is between Test 2 (Computation) and Test 6 (Arithmetic Reasoning). The lowest correlation (.06) is between Test 11 (Assemble) and Tests 6 and 8 (Mark Making). Test 7 (Form Perception) has high correlations with the tests of Clerical Perception (Test 1), Spatial Aptitude (Test),

and Form Perception (Test 5). With correlations of .51, .53, and .58, Test 6 (Arithmetic Reasoning) shows moderate relationships with measures of Clerical Perception, Spatial Perception and Vocabulary. Overall, the pencil and paper tests (Tests 1 to 8) tend to have higher intercorrelation coefficients than the apparatus board tests (Tests 9 to 12).

Table 11. GATB-ASL Correlation Matrix

Test No.	1	2	3	4	5	6	7	8	9	10	11	12
1	1.00											
2	.65	1.00										
3	.52	.50	1.00									
4	.48	.50	.45	1.00								
5	.55	.50	.53	.34	1.00							
6	.58	.75	.51	.53	.39	1.00						
7	.65	.52	.63	.45	.64	.44	1.00					
8	.59	.44	.35	.39	.44	.39	.44	1.00				
9	.43	.35	.31	.16	.26	.35	.35	.49	1.00			
10	.38	.22	.21	.12	.19	.16	.25	.53	.54	1.00		
11	.20	.07	.25	.21	.11	.06	.23	.06	.21	.11	1.00	
12	.26	.19	.22	.12	.22	.08	.27	.28	.37	.33	.48	1.00

The names of the twelve tests can be obtained from Table 9.

Extraction of Factors

The factor analysis program, SPSS for Windows 3.1, was used to extract from the correlation matrix factors with eigenvalues greater than one.

The unrotated factor matrix (Table 12) indicates three factors with eigenvalues greater than one. The first factor, with an eigenvalue of 5.232, accounts for approximately 48.3 per cent of the variance. The second factor, with an eigenvalue of 1.590, accounts for approximately 13.3 per cent of the variance. The third factor, with an eigenvalue of 1.207, accounts for approximately 10.1 per

cent of the variance. In total, the three factors account for approximately 66.9 per cent of the variance shared by the twelve tests of the battery. A Scree test for determining the number of factors to retain indicated three main factors, with a strong first factor.

Table 12. GATB-ASL Unrotated Factor Matrix

GATB-ASL No.	Name	Factor Loading F_1	F_2	F_3	Communality h^2
1	Name Comparison	.839	-.074	-.070	.715
2	Computation	.776	-.310	-.077	.705
3	Three-Dimensional Space	.727	-.157	.249	.616
4	Vocabulary	.629	-.306	.203	.532
5	Tool Matching	.701	-.166	.091	.528
6	Arithmetic Reasoning	.720	-.381	-.079	.670
7	Form Matching	.785	-.102	.182	.661
8	Mark Making	.707	.171	-.390	.682
9	Place	.597	.455	-.305	.657
10	Turn	.490	.544	-.465	.754
11	Assemble	.307	.477	.649	.805
12	Disassemble	.422	.651	.313	.701
Eigenvalue		5.232	1.590	1.207	Sum = 8.027
Percent of Variance		43.6	13.3	10.1	
Cumulative Percent		43.6	56.9	66.9	

Final Solution

Table 13 presents the three factor solution . With factor loadings from .674 to .787, Computation, Vocabulary, Arithmetic Reasoning, and Spatial Aptitude load heavily on Factor 1. Motor Coordination and Manual Dexterity load heavily on Factor 2, with factor loadings from .752 to .856. Finger Dexterity loads heavily on Factor 3, with loadings of .750 and .891.

All of the tests measuring Verbal, Numerical, Spatial, and Perceptual Aptitudes show a high loading on Factor 1 and a low loading on the other two factors. The tests which measure Manual Dexterity show a high loading on

Factor 2 and a low loading on the other two factors. Motor Coordination (Test 8) shows a high loading on Factor 2 and a low loading on the other two factors. Finger Dexterity shows a high loadings on Factor 3 and low loading on the other two factors.

Table 13. GATB-ASL Rotated (Orthogonal) Factor Matrix

GATB Tests		Factor Loadings		
Test No.	Test Name	F_1	F_2	F_3
2	Computation	**.787**	.290	-.035
6	Arithmetic Reasoning	**.782**	.223	-.096
7	Form Matching	**.722**	.238	.287
3	Three-Dimensional Space	**.718**	.133	.286
4	Vocabulary	**.716**	.031	.133
1	Name Comparison	**.705**	.447	.133
5	Tool Matching	**.674**	.219	.157
10	Turn	.004	**.856**	.129
9	Place	.168	**.762**	.221
8	Mark Making	.403	**.720**	.012
11	Assemble	.098	-.032	**.891**
12	Disassemble	.024	.371	**.750**

Note: Factor Loadings > .45 are in bold

Chapter 5

Discussion

In the following discussion, emphasis is placed on four main themes: the response of the test subjects to the instructions conveyed in American Sign Language (ASL), reliability estimates for the GATB-ASL, an exploratory factor analysis, and limitations to the study.

The study was undertaken primarily to determine whether or not deaf adults respond to instructions given in American Sign Language (ASL) in the manner intended by the developers of the General Aptitude Test Battery (GATB). Before answering the question, standard procedures for communicating the instructions of the GATB in ASL were developed. The development of the GATB-ASL involved translating the instructions into ASL and using animation to illustrate the correct responses to the items contained in the practice exercises. Standardization was achieved by presenting the instructions and the animation on videotape. Once the GATB-ASL had been developed, the performance of the hearing impaired sample was observed during administrations of the GATB-ASL.

As administrations of the GATB to hearing impaired persons modified the conditions with which the data on reliability and validity were obtained, one could not assume that the reliability and validity of the GATB also applied to the GATB-ASL. Therefore, it was necessary to acquire empirical evidence of the reliability and validity of the GATB-ASL.

Test-retest reliability was estimated with the Pearson Product Moment Correlation between scores obtained from the same test subjects on two occasions with an interval of six weeks between initial testing and retesting. Estimates of the internal consistency of the seven pencil and paper tests with multiple choice items were computed with Kuder-Richardson Formula Number 20 (KR-20). The KR-20 was not calculated for the five apparatus board tests as each result was the number of times single tasks were performed correctly.

In terms of validity, the intent was to identify the aptitudes measured by the GATB-ASL and then determine whether or not these aptitudes were similar to the aptitudes measured by the GATB. An exploratory factor analysis was used to extract factors with eigenvalues greater than one from a correlation matrix. The extracted factors were rotated to a simple structure, and the factorial structures of the GATB-ASL and the GATB were compared.

Response to Instructions

From a review of the literature, it was determined that most of the difficulties associated with the administration of tests to hearing impaired persons involve two main factors. One of the factors is the format and procedures of pencil and paper tests with multiple-choice items. The other factor is the English language reading grade level of test items.

Trybus (1973) noted that hearing impaired persons experienced difficulty with the tasks of matching items in test booklets with spaces on answer sheets and

filling-in the spaces on the answer sheets during timed tests. Although Trybus never made an attempt to explain his observation, Bragman's (1983b) study of the effects of different methods of conveying instructions on the performance of prelingual deaf children offered a plausible explanation. Bragman felt that most of the difficulties encountered in the administration of tests to deaf children could be attributed to the deaf child's lack of familiarity with the method of non-verbal communication that was used to convey instructions. Bragman noted that when deaf children were familiar with the method of communication and the format and procedures of testing, the children responded to instructions in an appropriately consistent manner. Bragman noted that regardless of the children's familiarity with the format and procedures of testing, their performance tended to be inappropriate and inconsistent when confronted with unfamiliar methods of communication. The inference to be drawn from Bragman's study is that deaf adults should be able to perform tasks in the manner intended by test developers when instructions are conveyed by a familiar method of communication.

The findings of the present study indicate that, test subjects experienced little difficulty performing the test taking tasks which Trybus (1973) described as being particularly difficult for hearing impaired persons. With the pencil and paper tests with multiple choice items, test subjects easily turned to the correct pages in the test booklets, placed the separate answer sheets next to the appropriate items in the test booklets, figured-out answers to arithmetic problems on scrap paper,

compared rows of slightly different shapes and figures, selected one of the multiple choices provided for each item, and filled-in the spaces on the answer sheets. From these observations, it was felt that the instructions conveyed in ASL provided sufficient clarity, emphasis, and detail so that test subjects were able to perform all of the test taking tasks in the manner intended by the developers of the GATB.

The other difficulty identified by Trybus (1973) was the English language reading grade level of test items. From a review of the literature, it was determined that a majority of deaf adults have an English language reading grade level of 4.0 (Morgan and Vernon, 1994). The obvious implication is that items that require a reading grade level above 4.0 may be too difficult for most deaf adults. Given that the United States Employment Service (1970) estimated the reading grade level of the GATB to be 6.0; the reading grade level of the GATB-ASL must also be 6.0. Therefore, it was felt that the reading grade level of Test 4 (Vocabulary), Test 6 (Arithmetic Reasoning) and the instructions printed at the bottom of the pages in the test booklets would be too difficult for most of the deaf adults in the sample.

Table 10 showed that the GATB-ASL Verbal and Numerical aptitude scores were approximately 1 to 1.5 standard deviations below the general working population mean of 100. Previous administrations of the GATB to hearing impaired persons (Botterbusch and Droege, 1972) also yielded Verbal and

Numerical aptitude scores that were approximately 1 to 1.5 standard deviations below the general working population mean of 100.

As the Verbal and Numerical aptitude scores of the GATB-ASL and the GATB represent the total number of correct responses to similar items, and as both sets of scores were recorded on a similar scale (a mean of 100 and a standard deviation of 20), the Verbal and Numerical aptitude scores of the GATB-ASL and the GATB appear to be very similar. One plausible explanation for the observed similarity would be that the reading grade level required by the tests which measure Verbal and Numerical aptitude may be above the reading grade level of most deaf adults. Such an assumption is congruent with Pintner's (1924) assertion that Intelligence tests which involve the use of the English language "immediately become subject matter tests for most deaf children". However, the reading grade level of the sample must be established before reading grade level can be shown to be the cause of low Verbal and Numerical scores.

The GATB-ASL scores for Perception and Motor Coordination fall within one standard deviation of the mean of the general working population. Botterbusch and Droege (1972) suggested that scores falling within one standard deviation of the mean of the general working population were an indication that deafness did not inhibit performance on measures of Perception and Motor Coordination. In other words, modified administrations of the GATB may have little observable effect on the reliability and validity of Perception and Motor Coordination. Botterbusch and Droege (1972), however, did not publish empirical evidence in

support of the reliability and validity of the measures yielded by modified administrations of the GATB.

Reliability Estimates

Estimates of internal consistency are high (.80 to .96) for Name Comparison (Test 1), Computation (Test 2), Three Dimensional Space (Test 3), Vocabulary (Test 4), Tool Matching (Test 5), Arithmetic Reasoning (Test 6) , and Form Matching (Test 7). The magnitude of the interitem consistencies yielded by the KR-20 suggests that each one of the items in each one of the pencil and paper tests has been taken from a highly homogeneous domain.

Test-retest reliability estimates tend to be high (.82 to .91) for Name Comparison (Test 1), Computation (Test 2), Three Dimensional Space (Test 3), and Mark Making (Test 8); moderate (.68 to .79) for Vocabulary (Test 4), Tool Matching (Test 5), Arithmetic Reasoning (Test 6), Place (Test 9), and Turn (Test 10); and low (.26 and .46) for Assemble (Test 11), and Disassemble (Test 12). Test retest reliability estimates indicate the extent to which scores can be generalized from initial testing to retesting, the higher the reliability the less susceptible the scores are to the effects of random error. In other words, the estimates of test retest reliability indicate the stability of the scores over time. Over a six week period, the stability of the scores yielded by the pencil and paper tests of the GATB-ASL appears to fluctuate from moderate and to high.

However, the stability of the tests that measure Finger Dexterity (Tests 11 and 12) appears to be relatively low.

Reliability estimates for the twelve tests of the GATB have never been reported, but reliability estimates for the aptitudes have been reported. Therefore, the test-retest reliability of the aptitudes measured by the GATB (Table 5) can be compared with the reliability of the aptitudes measured by the GATB-ASL (Table 10).

Overall, reliability estimates for the aptitudes measured by the GATB-ASL are not as high as the estimates for the aptitudes measured by the GATB. With a range of .67 to .95, the test retest reliability estimates for the GATB fluctuate from moderate to high. With a range of .40 to .92, the test retest reliability estimates for the GATB-ASL fluctuate from low to high. A comparison of the GATB with the GATB-ASL indicates that reliability estimates for Numerical, Spatial, Form Perception, Clerical Perception, Motor Coordination, and Manual Dexterity are very similar. A comparison of the reliability coefficients for Intelligence, Verbal, and Finger Dexterity shows that the GATB-ASL tends to be less stable over the six week interval than the GATB.

Finding a statistically significant increase in mean raw scores (Table 8) for Name Comparison, Computation, Three Dimensional Space, Tool Matching, Form Matching, Mark Making, Place, Turn, and Disassemble suggest that over a six week period scores on these tests may be susceptible to artificial inflation.

Such a finding is congruent with the National Research Council's (1989) report on the artificial inflation of GATB scores due to practice.

Exploratory Factor Analysis

With respect to construct validity, the United States Employment Service (1983) has suggested that the twelve tests of the GATB measure nine independent aptitudes which may be represented by three general factors. Verbal and Numerical Aptitudes may be represented by a Cognitive Factor. Spatial Aptitude and Form Perception may be represented by a Perceptual Factor. Motor Coordination, Manual Dexterity, and Finger Dexterity may be represented by a Psycho-Motor Factor.

If the GATB-ASL and the GATB measure the same aptitudes, then an exploratory factor analysis should reveal nine aptitudes which cluster in a way that is very similar to the three general factors identified by the USES (1983).

The exploratory factor analysis, suggests a simple orthogonal structure of three factors that represent 66.9% of the total variance. The tests measuring Verbal Aptitude, Numerical Aptitude, Spatial Aptitude, Clerical Perception, and Form Perception show a high loading on Factor 1 and a low loading on the other two factors. Tests measuring Motor Coordination and Manual Dexterity show a high loading on Factor 2. Tests measuring Finger Dexterity show a high loading on Factor 3 and a low loading on the other two factors.

It may be appropriate to describe the GATB-ASL Factor 1 as a general Cognitive Factor. A comparison of the GATB Factor 1 (Table 4) with the GATB-ASL Factor 1 (Table 13) indicates that Verbal, Numerical, Spatial, Form Perception, and Clerical Perception share medium to heavy loadings on Factor 1, and medium to light loadings on the other two factors. Therefore, the GATB-ASL Factor 1 is very similar to the GATB Factor 1 which has been described as a general Cognitive Factor.

The high loading of Mark Making on Factor 2 does raise a question about whether or not Factor 2 should be described as a measure of Manual Dexterity or Motor Coordination. The reliability of all three measures, as well as, their respective standard error of measurement fall within the same narrow range. Placing pegs in holes has the lowest reliability, mark making has the highest reliability, and turning pegs is in the middle. However, turning pegs loads highest of the three on Factor 2. Mark making is described as "the ability to coordinate the eyes and hands rapidly and accurately in making precise movements" (CEIC, 1986). Manual Dexterity is described as "the ability to move hands easily and skillfully" (USES, 1970). Given the lack of definitional clarity, it seems reasonable to suggest that, with respect to Factor 2, the descriptions of Motor Coordination and Manual Dexterity can be used interchangeably with little loss of meaning.

Factor 3 clearly represents Finger Dexterity which involves assembling small rivets and washers and disassembling them. The USES (1970) described Finger

Dexterity as "the ability to move the fingers and manipulate small objects with the fingers rapidly and accurately".

The factorial structures of the GATB-ASL and the GATB appear to be different. For example, the GATB-ASL three factor solution portrays Factors 2 and 3 as representing the Psycho Motor Aptitudes. Further, the first factor does not represent a clear differentiation between Perceptual, Verbal, and Numerical Aptitudes. Hence, the first factor may represent some general characteristic rather than a specific aptitude.

Limitations

To conduct random sampling, a population must be well defined and sample selection procedures must allow each member of the population an equal and independent opportunity to be selected. The population of hearing impaired people in Canada is not easily accessible which means any combination of recruitment methods inevitably results in some non-independence of the sample. As a consequence, the results of this study may not be generalizable to the population of deaf and hard-of-hearing adults.

The use of videotape as a means of standardizing the administration of signed instructions in a small group setting appears to be promising. However, measuring the effect of a video presentation independent of the difficulties associated with the multiple choice pencil-and-paper format is a practical problem which has not been resolved. As the scores yielded by tests with animated items

were similar to the scores yielded by tests with pictorial items that were not animated, the value of animation may be uniformity and ease of item presentation.

Conclusion

Overall, test subjects experienced little difficulty responding to the directions conveyed in American Sign Language. The performance of the test subjects indicate that most of the testing difficulties described by Trybus (1973) were very nearly eliminated by the conveying of the directions in a language with which the test subjects were familiar.

The English language reading grade level of items may have been too difficult for most of the test subjects, but that hypothesis can not be tested because the reading ability of the test subjects was never measured. Therefore, the validity of the inferences made from the measures of Verbal and Numerical Aptitude is beyond the scope of this study.

The high estimates of internal consistency suggest that the items in each of pencil-and-paper tests may have been drawn from highly homogeneous domains. With the exception of Finger Dexterity, the estimates of test-retest reliability suggest that the scores yielded by the GATB-ASL are very stable over a six week period. The low reliability estimates for Finger Dexterity suggest that these scores are not stable over a six week period.

The exploratory factor analysis orthogonally rotated three factors with eigenvalues greater than one to a simple structure. The first factor was interpreted as being a Cognitive Factor because the measures of Intelligence (G) and Perception (P, and Q) show a high loading on the first factor and a low loading on the other two factors. The second factor clearly represented Manual Dexterity. The third factor represented Finger Dexterity.

As test booklets, answer sheets, and apparatus boards of the GATB-ASL are identical to the standardized version of the GATB, the major deviation from the standard format involves presenting the directions on videotape. In presenting the practice exercises of Test 3, the videotaped presentation shows flat pieces of metal forming into three dimensional shapes which then move to become superimposed on the correct shapes. Although animation easily conveys the concept of bending, rolling, and folding two-dimensional shapes to form three-dimensional shapes, the effect of the videotape presentation on performance, independent of the format and procedures of pencil and paper tests with multiple choice items, remains unclear.

In closing, it seems appropriate to share a few of the practical insights which were acquired while conducting this study. It is the feeling of this researcher that multiple-choice items should never include "None of the above"; because, such a response runs the risk of slowing down deaf test takers which could place them at a disadvantage during tests with severe time limits. Further, the reading grade

level of each deaf test subjects should be determined prior to selecting and administering aptitude tests such as the GATB. Finally, non-signing test administrators should never assume that all deaf test subjects are equally fluent in American Sign Language.

Bibliography

Allen, T., White, C., and Karchmer, M. (1983). Issues in the development of a special edition for hearing impaired students of the 7th edition of the SAT. *American Annals of the Deaf,* 128, (34-39).

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1985). *Standards for Educational and Psychological Testing.* Washington, D.C: American Psychological Association.

American Psychological Association, Division 5: Evaluation Measurement and Statistics. (1993). Psychometric and assessment issues raised by the americans with disabilities act (ada). *The Score.* Washington, D. C: American Psychological Association

Anastasi, A. (1988). *Psychological Testing* (6th ed.). New York: MacMillan.

Anderson, R. and Sisco, F. (1977). *Standardization of the WISC-R performance scale for deaf children.* Washington, D. C: Gallaudet University, Center for Assessment and Demographic Studies. (Office of Demographic Studies Publication Series T, No. 1)

Bannowsky, A. W. (1983). Issues in assessing vocationally relevant personality factors prelingually deaf adults utilizing the 16 P. F. Form E. *Journal of Rehabilitation of the Deaf.* vol. 17, 3, (21-24)

Baydoun, R. B., and Neuman, G. A. (1992). The future of the general aptitude test battery (gatb) for use in public and private testing. *Journal of Business and Psychology.* vol. 7, no. 1, Fall.

Berger, D. G., Holdt, T. J., and Laforge, R. A. (1972). *Effective Vocational Guidance of the Adult Deaf.* Final report of the Oregon Vocational Research Project, Grant No. RD 2018-S. Rehabilitation Services Administration, Social Rehabilitation Service, Department of Health Education and Welfare, Washington, D.C.

Boatner, K. F., Stuckles, E.R., and Moores, D.F. (1964). *Occupational Status of the Young Adult Deaf of New England and the Need and Demand for a Regional Technical-Vocational Training Center.* West Hartford, Connecticut: American School for the Deaf.

Botterbusch and Droege. (1972). GATB aptitude testing of the deaf: problems and possibilities. *Journal of Employment Counselling*. vol.9. no.1, (14-19).

Bragman, R. (1982a). Review of research on test instructions for deaf children. *American Annals of the Deaf*. 127. (337-346)

Bragman, R. (1982b). Effects of different methods of conveying test directions on deaf children's performance on pattern recognition tasks. *Journal of Rehabilitation of the Deaf*. 16. (17-26)

Brown, F.G. (1983). *Principles of Educational and Psychological Testing* (3rd ed.). Toronto: Holt, Rinehart and Winston.

Butterworth, R. R. (1983). *The Perigee Visual Dictionary of Signing*. Toronto, ON: General Publishing Co.

Canada Employment and Immigration Commission (CEIC) (1986). *Manual for the General Aptitude Test Battery: Section I: Administration and Scoring*. Toronto: Nelson Canada

Canada Employment and Immigration Commission (CEIC) (1971). *Canadian Classification and Dictionary of Occupations: Vol. 1*. Ottawa, ON: Queen's Printer.

Canada Employment and Immigration Commission (CEIC) (1971). *Canadian Classification and Dictionary of Occupations: Vol. 2*. Ottawa, ON: Queen's Printer.

Carmines, E. G., and Zeller, R. A. (1979). *Reliability and Validity Assessment*. Newbury CA: Sage Publication (Series / No. 07-017).

Chovan, W., and Benfield, J. R. (1994). Varied sign language systems and their mediating effects on WISC -R verbal subtests of profoundly deaf students: a replication. *Perceptual and Motor Skills*. vol. 78, no.1 (61-62).

Clark, L. A., and Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychological Assessment*. vol. 7, no. 3 (309-319).

Cooney, D. J. (1994). *A Manual for the Administration of the General Aptitude Test Battery in American Sign Language*. Ottawa, ON: Algonquin College of Applied Arts and Technology

Crocker, L. and Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Toronto, Ont: Holt, Rinehart and Winston.

Darlington, R. B., and Carlson, P. M. (1987). *Behavioral Statistics: Logic and Method*. New York, NY: The Free Press.

Dwyer, C. L. and Wincenciak, S.L. (1977). A pilot investigation of three factors of the 16 P.F. Form E comparing the standard written form with an Ameslan videotape revision. *Journal of Rehabilitation of the Deaf.* vol,10. no.4, (17-23)

Floyd, F. J., and Widman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment.* vol. 7, no. 3 (286-299).

General Aptitude Test Battery, Form A. (1986). Toronto: Nelson Canada

Goetzinger, C. P., and Rousey, C. L. (1957). A study of weschler performance scale (form II) and the knox cube test with deaf adolescents. *American Annals of the Deaf.* vol.102, 329-339.

Greenberg, M., and Kusche, C. (1989). Cognitive, personal, and social development of deaf children and adolescents. In M. Wang, M. Reynolds, and H. Walberg (Eds.), *The Handbook of Special Education: Research and Practice.* (vol. 3., pp. 95- 129), Oxford England: Pergamon.

Gregory, R. J. (1992). *Psychological Testing: History, Principles, and Applications*. Toronto: Allyn and Bacon.

Growick, B., Kalliope, G., and Jones, Clair (1983). Sample norms for the hearing-impaired on select components of the valpar work sample series. *Vocational Evaluation and Work Adjustment Bulletin*. vol. 16, no. 2 (56-63).

Hartwig, F., and Dearing, B. E. (1979). *Exploratory Data Analysis*. Newbury CA: Sage Publications (Series / No. 07-016).

Hiskey, M. (1966). Hiskey-*Nebraska Test of Learning Aptitude: Manual*. Lincoln, NE: Union College Press.

Hoeman, H.W. (1972). Communication accuracy in a sign language interpretation of a group test. *Journal of Rehabilitation of the Deaf.* vol.5, (40-43).

Holm, C. S. (1987). Testing for values with the deaf: the language/ cultural effect. *Journal of Rehabilitation of the Deaf.* 20. 4. (7- 17).

Jensema, C. (1975). Reliability of the 16PF Form E for hearing impaired college students. *Journal of Rehabilitation of the Deaf.* vol.8,no.3, (14-18)

Jacobs, L. R. (1977). The efficiency of interpreting input for processing lecture information by deaf college students. *Journal of Rehabilitation of the Deaf.* vol. 11, no. 2 (10-11).

Kronenberg, H. H., and Blake, G. D. (1966). *Young Deaf Adults: An Occupational Survey.* Washington, D. C: Vocational rehabilitation Administration.

Kim, J., and Mueller, C. W. (1978). *Introduction to Factor Analysis.* Newbury Park, CA: Sage Publications (Series / No. 07-013)

Kim, J. and Muller, C. W. (1978). *Factor Analysis: Statistical Methods and Practical Issues.* Newbury CA: Sage Publications (Series / No. 07-014)

Luckner, J. L. and McNeil, J. H. (1994). Performance of a group of deaf and hard - of - hearing students and a comparison group of hearing students on a series of problem - solving tasks. *American Annals of the Deaf.* vol.139, no.3, (371-377).

Levine, E. S. (1974). Psychological tests and practices with the deaf: a survey of the state of the art. *Volta Review*, 76. (298-319).

Messick, S. (1988). The once and future issues of validity: assessing the meaning and consequences of measurement. In H. Wainer and H. Braun (Eds.), *Test Validity.* (pp. 33 - 47). Hillsdale, N.J: Lawrence Erlbaum Associates.

Miller, M. S. (1984). Experimental use of signed presentation of the Verbal scale of the WISC-R with profoundly deaf children: a preliminary report of the sign selection process and experimental test procedures. In D. S. Martin (Ed.), *International Symposium on Cognition, Education, and Deafness: Working papers; I and II.* Washington, D.C: Gallaudet College (ERIC Document Reproduction Service No. ED 247 727)

Moores, D., Harlow, M. J., and Fisher, S. (1974). Current research, evaluation of post secondary programs for the deaf (Personal Correspondence with Robert Sanderson). In Robert Sanderson, *The effects of educational Backgrounds of Deaf Children on their General Aptitude Test battery Performance Scores. Doctoral Dissertation, Education, Guidance, and Counseling. Brigham Young University.*

Morgan, A. and Vernon, M. (1994). A guide to the diagnosis of learning disabilities in deaf and hard of hearing children and adults. *American Annals of the Deaf.* vol. 139, no. 3, (358-370).

Murphy, J. H. and Fleischer, L. R. (1977). The effects of ameslan versus siglish upon test scores. *Journal of Rehabilitation of the Deaf.* 11. 2. (15-18)

National Research Council (NRC). (1989). *Fairness in Employment Testing: Validity Generalization, Minority Issues, and the General Aptitude Test Battery.* (J. A. Hartigan and Alexander K. Wigdor, Eds.). Washington, D. C: National Academy Press.

Nelson Canada. (1986). General Aptitude Test Battery, Form A: Scoring Template. Toronto, ON: Nelson Canada

Norris, S. P. (1995). Format effects on critical thinking test performance. *The Alberta Journal of Educational Research.* vol. XLI, no. 4 (378- 406)

Paul, P. and Jackson, D. W. (1993). *Toward a Psychology of Deafness: Theoretical and Empirical Perspectives.* Toronto: Allyn and Bacon.

Pintner, R. (1924). *Intelligence Testing.* New York: Henry Holt and Company

Psychological Corporation. (1956). How accurate is a test score. *Test Service Bulletin.* June, no. 50

Reynolds, W. M., and Kobak, K. A. (1995). Reliability and validity of the hamilton depression inventory: a paper-and-pencil version of the hamilton depression rating scale clinical interview. *Psychological Assessment.* vol. 7., no. 4 (472-483)

Rodda, M. and Grove, C. (1987). *Language, Cognition and* Deafness. Hillsdale, New Jersey: Lawrence Erlbaum Associates

Rosen, A. (1967). Limitations of of personality inventories for assessment of deaf children and adults as illustrated by research with the MMPI. *Journal of Rehabilitation of the Deaf.* 1. (47-52)

Rubin, D. (1988). Discussion. In H. Wainer and H. Braun (Eds.), *Test Validity*. (pp. 241 - 256). Hillsdale, N.J: Lawrence Erlbaum Associates.

Salvia, J. and Ysseldyke, J. E. (1991). *Assessment* (5th ed.). Boston: Houghton Mifflin Co.

Sanderson, R.G. (1974). *The Effect of Educational Background of Deaf Children on Their General Aptitude Test Battery Performance Score*. (unpublished doctoral dissertation). Brigham Young University. University Microfilms, A Xerox Company, Ann Arbour Michigan. File no. 74-18, 244

Sattler, J. (1992). *Assessment of Children* (revised and updated 3rd edition). San Diago: Sattler Publications.

Schein, J. D. (1984). *Speaking the Language of Sign: The Art and Science of Signing*. Garden City, NY: Doubleday

Schein, J. D. (1992). *Canadians with Impaired Hearing: The Health and Activities Limitation Survey*. (catalogue No. 82-615, Vol. 5). Ottawa, ON: Statistics Canada.

Shiels, J.W. (1980). Work samples and normative data: are deaf norms really needed. *Journal of Rehabilitation of the Deaf*. vol. 14, no.1 (14-21)

Sherman, E. M. S., Strauss, E., Spellacy, F. Hunter, M. (1995). Construct validity of WAIS-R factors: Neuropsychological test correlates in adults referred for evaluation of possible head injury. *Psychological Assessment*. vol. 7, no. 4 (440-444).

Shroyer, E. H. and Shroyer, S. P. (1984). *Signs Across America*. Washington, D. C: Gallaudet College Press.

Spidal, D. A. and Pfau, G. S. (1972). The potential for language acquisition of illiterate deaf adolescents and adults. *Journal of rehabilitation of the Deaf*. vol. 6. 1, (27- 33).

Sullivan, P. M., and Vernon, M. (1979). Psychological assessment of hearing impaired children. *School Psychology Digest*. 8, (271-289).

Traub, R. E. (1994). *Reliability for the Social Sciences: Theory and Applications*. Thousand Oaks: Cal. Sage Publication

Traxler, C. B. (1989). The role of assessment in placing deaf students in academic and vocational courses. In T. E. Allen, B. W. Rawlings, and A. N. Schildroth, (Eds.), *Deaf Students in the School to Work Transition.* (pp. 141-187). Toronto: Paul H Brooks

Trybus, R. J. (1973). Personality assessment of entering hearing-impaired college students using the 16 P. F., form E. *Journal of Rehabilitation of the Deaf.* vol.3, no. 6 (34-40)

Tweney, R. D., and Hoeman, H.W. (1976). Translation and sign language. In R. W. Brislin (Ed.), T*ranslation: Applications and Research.* (pp. 138-161) New York: Gardner Press.

United States Employment Service (USES). (1970). *Manual for the USES General Aptitude Test Battery: Section III: Development.* Washington, D. C: United States Department of Labor, Manpower Administration. U. S. Government Printing Office: Stock No. 029-014-00018-3.

United States Employment Service (USES). (1983). *The Dimensionality of the General Aptitude Test Battery (GATB) and the Dominance of General Factors over Specific Factors in the Prediction of Job Performance.* (USES Test Research Report No. 44). Washington D. C: Division of Counselling and Test Development, Employment Training Administration, U. S. Department of Labor.

Vernon, M. (1968). Fifty years of research on the intelligence of deaf and hard of hearing children: a review of literature and discussion of implications. *Journal of Rehabilitation of the Deaf.* vol.1, (1-12).

VISTA. (1988). American Sign Language Series:*Signing Naturally (video).* San Diego, CA: Dawn Sign Press

Watson, D. (1979). Guidelines for the psychological and vocational assessment of deaf rehabilitation clients. *Journal of Rehabilitation of the Deaf.* vol.13, (27-57)

Willingham, W. W. (1988). Testing handicapped people- the validity issue. In H. Wainer and H. Braun (Eds.), *Test Validity.* (pp. 89 -105). Hillsdale, N.J: Lawrence Erlbaum Associates.

Appendix A

The General Aptitude Test Battery

Descriptions of the Twelve Tests of the Battery

The descriptions of the twelve tests were taken from the United States Employment Service. (1970). *Manual for the USES General Aptitude Test Battery, Section III: Development.* Washington, D. C: United States Department of Labor.

1. **Name Comparison** (150 items; time limit, 6 minutes)
   The test consists of two columns of names. The examinee inspects each pair of names, one in each column, and indicates whether the names are the same or different.

2. **Computation** (50 items; time limit, 6 minutes)
   The test consists of a number of arithmetic exercises requiring the addition, subtraction. multiplication, or division of whole numbers.

3. **Three Dimensional Space** (40 items, time limit, 6 minutes)
   The test consists of a series of exercises containing a stimulus figure and four drawings of three-dimensional objects. The stimulus figure is pictured as a flat piece of metal which is to be either bent, or rolled, or both. Dotted lines indicate where the stimulus figure is to be bent. The examinee indicates which one of the four drawings of three-dimensional objects can be made from the stimulus figure.

4. **Vocabulary** (60 items, time limit, 6 minutes)
   The test consists of sets of four words. The examinee indicates which two words have either the same or opposite meanings.

5. **Tool Matching** (49 items, time limit, 5 minutes)
   The test consists of a stimulus drawing and four black-and-white drawings of simple shop tools. The examinee indicates which of the four drawings is the same as the stimulus drawing. Variation exist only in the distribution of black and white in each drawing.

6. **Arithmetic Reasoning** (25 items, time limit, 7 minutes)
   The test consists of a number of arithmetic problems expressed verbally.

7. **Form Matching** (60 items, time limit, 6 minutes)

   The test consists of two groups of variously shaped line drawings. The examinee indicates which figure in the second group is exactly the same size and shape as each figure in the first or stimulus group.

8. **Mark Making** (130 items; time limit, 60 seconds)

   The test consists of a series of squares in which the examinee is to make three pencil marks, working as rapidly as possible. The marks to be made are short lines, two vertical and the third a horizontal line beneath them.

9. **Place** (144 items; total time limit, 45 seconds)

   The equipment used for this test and for Part 10 consists of a rectangular pegboard divided into two sections, each section containing 48 holes. The upper section contains 48 cylindrical pegs. The examinee removes the pegs from the holes in the upper part of the board and inserts them in the corresponding holes in the lower part of the board, moving two pegs simultaneously, one in each hand. This performance is done three times, with examinees working rapidly to move as many of the pegs as possible during the time allowed for each of the three trials.

10. **Turn** (144 items, total time limit, 90 seconds)

    For Part 10, the lower section of the board described under Part 9 contains the cylindrical pegs, The examinee removes a peg from a hole, turns the peg over so that the opposite end is up, and returns the peg to the hole from which it was taken, using only the preferred hand. The examinee works rapidly to turn and replace as many of the 48 pegs as possible during the time allowed. Three trials are given for this performance.

11. **Assemble** (50 items, time limit, 90 seconds)

    The equipment used for this test and for Part 12 consists of a small rectangular board containing 50 holes, and a supply of small metal rivets and washers. The examinee takes a small metal rivet from a hole in the upper part of the board with the preferred hand and at the same time removes a small metal washer from a vertical rod with the other hand; examinee puts the washer on the rivet, and inserts the assembled piece into the corresponding hole in the lower part of the board using only the preferred hand. The examinee works rapidly to assemble and place as many rivets and washers as possible during the time allowed.

12. **Disassemble** (50 items, time limit, 60 seconds)
The examinee removes the small metal rivet of the assembly from a hole in the lower part of the board, slides the washer to the bottom of the board, puts the washer on the rod with one hand and the rivet into the corresponding hole in the upper part of the board with the other (preferred) hand. The examinee works rapidly to disassemble and replace as many rivets and washers as possible during the time allowed.

Descriptions of the Nine Aptitudes Measured by the GATB

Descriptions of the aptitudes were taken from Employment and Immigration Canada. (1986). *Manual for the Administration and Scoring of the General Aptitude Test Battery.* Toronto, ON: Nelson Canada.

(G) General Learning Ability

General Learning Ability is the ability to "catch-on" or understand instructions and underlying principles, the ability to reason and make judgments. Closely related to doing well in school.

(V) Verbal Aptitude

The ability to understand meaning of words and to use them effectively. the ability to comprehend language, to understand relationships between words and to understand meanings of whole sentences and paragraphs.

(N) Numerical Aptitude

The ability to perform arithmetic operations quickly and accurately.

(S) Spatial Aptitude

The ability to think visually of geometric forms and to comprehend the two-dimensional representation of three dimensional objects. The ability to recognize the relationships resulting from the movement of objects in space.

(P) Form Perception

The ability to perceive pertinent detail in objects or in pictorial or graphic material. The ability to make visual comparisons and discriminations and see slight differences in shapes and shadings and widths and lengths of lines.

(Q) Clerical Perception

The ability to perceive pertinent detail in verbal or tabular material. The ability to observe differences in copy, to proofread words and numbers, and to avoid perceptual errors in arithmetic computation. A measure of speed of perception which is required in many industrial jobs even when the job does not have verbal or numerical content.

(K) Motor Coordination

The ability to coordinate eyes and hands or fingers rapidly and accurately in making precise movements with speed. The ability to make a movement response accurately and swiftly.

(F) Finger Dexterity

The ability to move the fingers, and manipulate small objects with the fingers rapidly or accurately.

(M) Manual Dexterity

The ability to move the hands easily and skillfully. the ability to work with the hands in placing and turning motions.

Appendix B

GATB-ASL: Typical Questions and Problems

The following observations were originally presented in: Cooney, D. (1994). *Manual for the Administration of the General Aptitude Test Battery in American Sign Language*. Ottawa, ON: Algonquin College of Applied Arts and Technology.

Typical Questions

The following section is not an inclusive list by any means. However, care was taken to record the most frequently asked questions. As all of the questions recorded here concern the procedures of the pencil and paper tests with multiple choice items, the presentation is organized so that questions are given in the order with which they were asked.

Test 1, Name Comparison

Q. Do we fill-in the answer bubble completely or can we just mark an "X" in the bubble?

A. Let me see your answer sheet, please.

(look to see if the answer mark shows through the scoring template and gives every indication of being a clearly understood answer)

Give a reasonable response to the individual.

Q. Do we have to do every question in six minutes? That's impossible!

A. No, you do not do every question in six minutes. You do as many as you can in six minutes. The test was set-up so everyone would have a different number of answers, your job is to work quickly and answer as many questions as you can in six minutes.

Test 2, Computation

Q. What does that mark (division sign) mean?

A. It means to divide.

Q. I was never taught how to divide?

A. Skip those questions and do the ones you can do.

Test 4, Vocabulary

Q. Are the words the same or are the words different? I don't understand !

A. In every question there will be two words that are the same or there will be two words that are different. You have to decide if you are looking for words that are the same or if you are looking for words that are different. In every question there will be only one choice, you decide if two words are the same or different.

Test 5, Tool Matching

Q. Why does the video keep repeating the same directions?

We understand, this is easy!

A. Hearing people are given the same directions every time. these directions were set-up a long time ago for hearing people and we have not changed the directions for deaf people. You know, these are easy directions, but some people still write in the test books. So please be patient and follow the directions.

Test 6, Arithmetic Reasoning

Q. May I use my calculator?

A. No, when the GATB was first developed calculators were not used; so we do not know what your score would mean if you were allowed to use calculators. Maybe, later, in the future, calculators will be allowed, but for today we will not use calculators.

It is important to note that the body language of the test administrator can be a more important consideration for achieving standardization than responding to questions in a consistent manner. The way the test administrator moves, looks, and signs when answering questions can be as significant to the deaf person as the actual answer. Regardless of the question, it is important to respond to each person with body language that communicates respect, genuineness, patience, and generosity of spirit.

Typical Problems

Test 4, Vocabulary

The most common problem was confusion over the meaning of the words "amusing" and "funny" in practice exercise number five. The most effective explanation appeared to be pointing to "amusing" and signing "happy"; and pointing to "funny" and signing "happy"; then pointing to both "amusing" and "funny" and signing "same".

Test 5, Tool Matching

By the time test subject began Test 5, some complained that the test is too easy and begin to ignore the video instructions. The test administrator used rapport to gently encourage these participants to continue watching the video.

A typical problem was jumping ahead. One or two participants began doing Test 5 before being told to begin. To indicate clearly the jumping ahead is unacceptable behavior without disrupting the other participants, the test administrator calmly took-away the individual's answer sheet and gave in-return a blank answer sheet already opened to Test 5. At the end of Test 5, the original answer sheet was returned to the individual so that person could continue working with the original answer sheet.

Test 9, Place

During Test 9, the most common error was made by moving the pegs from the lower part of the upper board to the upper part of the lower board.

Test 10, Turn

The most common error was made by turning the pegs from right to left for the first row, then turning the pegs from left to right for the second row.

Test 11. Assemble

The most common error was made by putting washers on the lower board then moving the washer with the rivet before placing the rivet in the upper hole of the first column on the lower part of the board.

Test 12. Disassemble

During Test 12, the most common error was made by moving the washer to the side of the board before picking-up the washer.

In many cases, the test administrator learned to anticipate which test subjects were about to make any one of the above errors before it occurred. Attempts to correct the error before it occurred confused the test subject and frustrated the test administrator. The best strategy for correcting the test subject appeared to be waiting for the error to be made. Hold the test subjects hands for a few seconds to stop further action, then use pointing and pantomime to demonstrate the appropriate procedure. In every case, pointing and pantomime resulted in the test subject recognizing the error and acting correctly. After being corrected, a few test subjects would repeat the exercise from the beginning. If it was important to make a correction, then it is important to allow the test subject the opportunity to repeat the entire practice exercise.

Appendix C

Computer Print-out of the Factor Analysis

```
01 Feb 96 SPSS for MS WINDOWS Release 6.1

This software is functional through January 31, 1997.


Data written to the working file.
14 variables and 106 cases written.
Variable: NO          Type: Number   Format: F8.2
Variable: ID          Type: Number   Format: F8.2
Variable: V1          Type: Number   Format: F8.2
Variable: V2          Type: Number   Format: F8.2
Variable: V3          Type: Number   Format: F8.2
Variable: V4          Type: Number   Format: F8.2
Variable: V5          Type: Number   Format: F8.2
Variable: V6          Type: Number   Format: F8.2
Variable: V7          Type: Number   Format: F8.2
Variable: V8          Type: Number   Format: F8.2
Variable: V9          Type: Number   Format: F8.2
Variable: V10         Type: Number   Format: F8.2
Variable: V11         Type: Number   Format: F8.2
Variable: V12         Type: Number   Format: F8.2
```

- - - - - - - - - - - FACTOR ANALYSIS - - - - - - - - - - -

Analysis number 1 Listwise deletion of cases with missing values

	Mean	Std Dev	Label
V1	90.11321	19.74375	
V10	72.58491	10.66268	
V11	36.19811	20.75867	
V12	54.33962	11.21554	
V2	58.33019	18.59171	
V3	97.83019	21.32982	
V4	62.00943	17.71171	
V5	51.94340	11.77878	
V6	14.00000	4.22464	
V7	46.12264	11.65267	
V8	102.05660	21.29468	
V9	22.63208	18.38546	

Number of Cases = 106

Correlation Matrix:

	V1	V10	V11	V12	V2	V3	V4
V1	1.00000						
V10	.38019	1.00000					
V11	.20311	.11986	1.00000				
V12	.26128	.33926	.48764	1.00000			
V2	.65224	.22727	.07092	.19467	1.00000		
V3	.52075	.21941	.25582	.22096	.50086	1.00000	
V4	.48491	.12272	.21551	.12349	.50185	.45382	1.00000
V5	.55792	.19932	.11935	.22515	.50831	.53828	.34298
V6	.58563	.16914	.06071	.08945	.75481	.51334	.53674
V7	.65370	.25758	.23940	.27149	.52589	.63646	.45831
V8	.59236	.53116	.06435	.28560	.44956	.35285	.39058
V9	.43016	.54240	.21569	.37385	.35672	.31737	.16122

	V5	V6	V7	V8	V9
V5	1.00000				
V6	.39771	1.00000			
V7	.64918	.44380	1.00000		
V8	.44570	.39604	.44949	1.00000	
V9	.26364	.35497	.35233	.49275	1.00000

- - - - - - - - - - FACTOR ANALYSIS - - - - - - - - - -

1-tailed Significance of Correlation Matrix:

' . ' is printed for diagonal elements.

	V1	V10	V11	V12	V2
V1	.				
V10	.00003	.			
V11	.01839	.11051	.		
V12	.00341	.00019	.00000	.	
V2	.00000	.00957	.23501	.02277	.
V3	.00000	.01192	.00406	.01142	.00000
V4	.00000	.10507	.01326	.10363	.00000
V5	.00000	.02026	.11150	.01016	.00000
V6	.00000	.04152	.26823	.18093	.00000
V7	.00000	.00384	.00673	.00244	.00000
V8	.00000	.00000	.25612	.00150	.00000
V9	.00000	.00000	.01319	.00004	.00009

	V3	V4	V5	V6	V7
V3	.				
V4	.00000	.			
V5	.00000	.00016	.		
V6	.00000	.00000	.00001	.	
V7	.00000	.00000	.00000	.00000	.
V8	.00010	.00002	.00000	.00001	.00000
V9	.00046	.04937	.00316	.00009	.00011

	V8	V9
V8	.	
V9	.00000	.

Extraction 1 for analysis 1, Principal Components Analysis (PC)

---------- FACTOR ANALYSIS ----------

Initial Statistics:

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
V1	1.00000	*	1	5.23276	43.6	43.6
V10	1.00000	*	2	1.59061	13.3	56.9
V11	1.00000	*	3	1.20718	10.1	66.9
V12	1.00000	*	4	.81733	6.8	73.7
V2	1.00000	*	5	.62750	5.2	79.0
V3	1.00000	*	6	.55772	4.6	83.6
V4	1.00000	*	7	.43465	3.6	87.2
V5	1.00000	*	8	.40020	3.3	90.6
V6	1.00000	*	9	.34474	2.9	93.4
V7	1.00000	*	10	.33182	2.8	96.2
V8	1.00000	*	11	.24244	2.0	98.2
V9	1.00000	*	12	.21305	1.8	100.0

---------- FACTOR ANALYSIS ----------

Hi-Res Chart # 1:Factor scree plot

PC extracted 3 factors.

Factor Matrix:

	Factor 1	Factor 2	Factor 3
V1	.83941	-.07435	-.07038
V10	.49055	.54477	-.46545
V11	.30757	.47700	.69491
V12	.42205	.65198	.31368
V2	.77679	-.31011	-.07751
V3	.72726	-.15720	.24985
V4	.62975	-.30664	.20392
V5	.70158	-.16622	.09112
V6	.72025	-.38149	-.07935
V7	.78597	-.10222	.18283
V8	.70765	.17120	-.39081
V9	.59700	.45593	-.30580

Final Statistics:

Variable	Communality	*	Factor	Eigenvalue	Pct of Var	Cum Pct
V1	.71508	*	1	5.23276	43.6	43.6
V10	.75405	*	2	1.59061	13.3	56.9
V11	.80503	*	3	1.20718	10.1	66.9
V12	.70161	*				
V2	.70557	*				
V3	.61605	*				
V4	.53220	*				
V5	.52815	*				
V6	.67059	*				
V7	.66162	*				
V8	.68281	*				
V9	.65779	*				

- - - - - - - - - - - - FACTOR ANALYSIS - - - - - - - - - - - -

EQUAMAX rotation 1 for extraction 1 in analysis 1 - Kaiser Normalization.

EQUAMAX converged in 6 iterations.

Rotated Factor Matrix:

	Factor 1	Factor 2	Factor 3
V1	.70529	.44711	.13323
V10	.00410	.85869	.12915
V11	.09892	-.03285	.89116
V12	.02490	.37176	.75019
V2	.78736	.29055	-.03503
V3	.71823	.13373	.28690
V4	.71646	.03114	.13385
V5	.67483	.21906	.15739
V6	.78202	.22317	-.09607
V7	.72253	.23834	.28770
V8	.40377	.72084	.01288
V9	.16816	.76200	.22106

Factor Transformation Matrix:

	Factor 1	Factor 2	Factor 3
Factor 1	.80470	.52604	.27520
Factor 2	-.56715	.54413	.61827
Factor 3	.17549	-.65361	.73621

Bi-Res Chart # 2:Factor plot of factors 1, 2, 3

Factor Score Coefficient Matrix:

	Factor 1	Factor 2	Factor 3
V1	.14536	.09705	-.02767
V10	-.18647	.48768	-.04631
V11	-.02176	-.18215	.62538
V12	-.12197	.09562	.46693

- - - - - - - - - - - FACTOR ANALYSIS - - - - - - - - - - -

	Factor 1	Factor 2	Factor 3
V2	.21876	.01397	-.12696
V3	.20421	-.11594	.12952
V4	.23583	-.15200	.03829
V5	.18040	-.03567	.02786
V6	.23525	-.01513	-.15880
V7	.18389	-.05494	.11310
V8	-.00903	.34130	-.13458
V9	-.11522	.38155	.02213

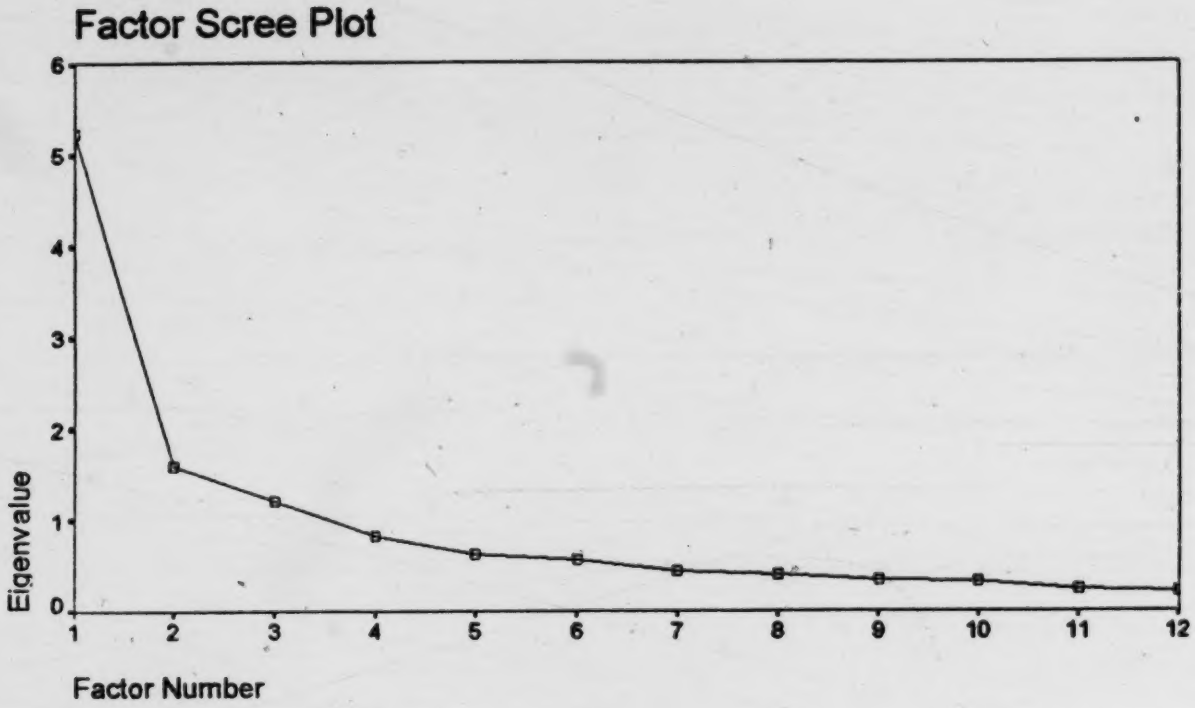
Covariance Matrix for Estimated Regression Factor Scores:

	Factor 1	Factor 2	Factor 3
Factor 1	1.00000		
Factor 2	.00000	1.00000	
Factor 3	.00000	.00000	1.00000

3 PC EXACT factor scores will be saved.

Following factor scores will be added to the working file:

Name	Label
FAC1_1	REGR factor score 1 for analysis 1
FAC2_1	REGR factor score 2 for analysis 1
FAC3_1	REGR factor score 3 for analysis 1


Factor Scree Plot
Eigenvalue
0
1
2
3
4
5
6
1
2
3
4
5
6
7
8
9
10
11
12
Factor Number

Appendix D

Questionnaire

QUESTIONNAIRE

(CONFIDENTIAL WHEN COMPLETED)

PLEASE PRINT

1. _____ _____
   First Name Last Name

2. _____ Date of Birth: _____
   Age Day Month Year

3. What is your highest completed level of formal education?

   _____

4. I'm not sure but I was about _____ when I became deaf.
   Years of age

5. Please tell how well you hear, by marking one of the statements below for each ear.

In my **LEFT** ear		In my **RIGHT** ear	
☐	I am deaf	☐	I am deaf
☐	I have a lot of trouble hearing	☐	I have a lot of trouble hearing
☐	I have a little trouble hearing	☐	I have a little trouble hearing
☐	My hearing is good	☐	My hearing is good

6. I have a job right now. ☐ Yes ☐ No

7. My job is full-time. ☐ Yes ☐ No

8. My job is part-time. ☐ Yes ☐ No

9. My job is _____

10. I have not been working for _____
How long

11. My last job was full-time. ☐ Yes ☐ No

12. My last job was part-time. ☐ Yes ☐ No

13. My last job was _____

14. I was in my last job for _____
How long

FINISHED. THANK YOU.

PLEASE CHECK TO ENSURE THAT YOU ANSWERED ALL QUESTIONS ACCURATELY.

CONFIDENTIAL WHEN COMPLETE

NOTES: _____

_____

_____

_____